PUBLIC (PUBLIQUE)
SAP Predictive Analytics
2023-12-14

Aide en ligne Automated Analytics

© 2023 SAP SE ou société affiliée SAP. Tous droits réservés.

THE BEST RUN SAP

Contenu

1 Nouveautés de l'aide Automated Analytics. 7
1.1 Historique du document. 7

2 Où trouver des informations complémentaires ?. 10
2.1 Ce que vous devez savoir avant de commencer à utiliser SAP Predictive Analytics. 11
2.2 Ce que vous devez savoir en tant qu'administrateur. 12
2.3 Ce que vous devez savoir en tant qu'analyste. 13

3 Panneau de démarrage. 15
3.1 Présentation de SAP Predictive Analytics. 15
3.2 À propos de Data Manager. 16
3.3 À propos de Automated Analytics. 17
3.4 À propos d'Expert Analytics. 18
3.5 À propos de Toolkit. 19

4 Data Manager. 20
4.1 Connexion à un référentiel d'objets analytiques. 20
4.2 Définition de la nouvelle manipulation de données. 20
4.3 Gestion des données analytiques. 21
Définitions des objets. 21
Actions sur les objets. 23
Options d'affichage. 23
4.4 Éditeur d'enregistrements analytiques. 24
Accueil. 25
Vues. 31
Édition. 38
4.5 Éditeur d'expressions. 39
4.6 Définition d'un agrégat. 45
Paramètres d'agrégation. 45
Paramètres de période. 46
Filtres et pivots. 47
4.7 Nouvelle condition. 49
4.8 Ajout d'une nouvelle table de recherche. 49
4.9 Nouvelle normalisation. 50
4.10 Nouvelles fonctions. 51
4.11 Nouvelles expressions SQL. 52
4.12 Modification de jeu de données d'analyse temporelle. 52
4.13 Modification d'un enregistrement analytique. 53

4.14 Modification d'une entité. . . . 55
4.15 Modification des filtres. . . . 56
4.16 Modification des filtres. . . . 61
4.17 Agrégats glissants. . . . 61
4.18 Composition de populations estampillées. . . . 61
4.19 Produit cartésien : population estampillée. . . . 62
4.20 Population estampillée filtrée. . . . 64
4.21 Produit cartésien. . . . 65
4.22 Ajout d'un filtre dans un jeu de données. . . . 66
Ajout d'une condition. . . . 66
Ajout d'une conjonction logique. . . . 67
Changement de l'ordre. . . . 68
Suppression d'un nœud. . . . 68
Aperçu des données. . . . 68
Enregistrement du filtre. . . . 68
Chargement d'un filtre. . . . 69
Échantillonnage. . . . 69
4.23 Source des données d'événement. . . . 70
Jeu de données d'événement pour la vérification des écarts. . . . 70
Jeu de données d'événements pour la simulation. . . . 71
4.24 Description des données d'événement. . . . 72
Définitions. . . . 72
Description des données. . . . 74
Enregistrement d'une description. . . . 75
4.25 Application de données événementielles. . . . 76
4.26 Gestion des performances lors de l'utilisation des bases de données. . . . 76

5 Automated Analytics Modeler. . . . 79
5.1 Ouverture ou suppression d'un modèle. . . . 79
Ouverture ou suppression d'un modèle. . . . 79
Ouverture ou suppression d'une manipulation de données. . . . 81
Ouverture d'un modèle Social. . . . 82
5.2 Ajouter une fonctionnalité de modélisation. . . . 83
5.3 Sélection des données. . . . 84
Sources de données prises en charge. . . . 84
Données à modéliser. . . . 87
Description des données. . . . 91
Paramètres avancés de l'espace. . . . 106
Stratégie de partition. . . . 109
Aperçu des données. . . . 110
5.4 Définition des paramètres. . . . 113
Sélection de variables. . . . 113

Récapitulatif des paramètres de modélisation. 115
Définition de paramètres pour la classification/régression. 118
Définition de paramètres pour la segmentation/mise en cluster. 126
Définition de paramètres pour les séries temporelles. 130
Définition de paramètres pour les règles d'association. 136
Définition des paramètres pour la journalisation d'événements. 139
Définition de paramètres pour l'analyse de séquences. 145
Définition de paramètres pour l'analyse de texte. 150
5.5 Apprentissage d'un modèle. 153
Apprentissage du modèle. 153
Encodage textuel - Apprentissage. 162
Vérification du modèle. 162
5.6 Utilisation du modèle. 163
Classification/Régression. 163
Mise en cluster. 164
Série chronologique. 165
Règles d'association. 166
Options autonomes de journalisation des événements et d'analyseur de séquences. 166
Options d'encodage textuel seul. 167
Mode Sans application ou Simulation. 168
Analyse de la vérification des écarts. 170
5.7 Affichage d'un modèle. 171
Affichage de la classification/régression. 171
Affichage de la segmentation/mise en cluster. 187
Exploration des règles d'association. 189
Affichage des séries temporelles. 193
Paramètres du modèle. 199
Options de rapports statistiques. 199
Rapports. 200
Personnalisation des feuilles de style de rapports. 200
5.8 Exécution du modèle. 202
Application d'un modèle. 202
Simulation du modèle. 212
Analyse des écarts. 213
Analyse des écarts. 214
Sélection des variables contributives. 215
5.9 Enregistrement/Export. 217
Génération du code. 217
Génération d'un script KxShell. 220
Enregistrement du modèle. 222
Enregistrement et exportation d'un jeu de données d'analyse. 223

Sauvegarde automatique. . . . 224
5.10 Social. . . . 225
Concepts. . . . 225
Définition des paramètres. . . . 235
Utilisation du modèle. . . . 250
Géolocalisation : Co-emplacement et chemin. . . . 260
5.11 Recommandation. . . . 280
Définition des paramètres. . . . 280
Utilisation du modèle. . . . 281
5.12 Calcul. . . . 292

6 Expert Analytics. . . . 293
6.1 Aide en ligne d'Expert Analytics. . . . 293

7 Toolkit. . . . 294
7.1 Explorateur de jeux de données. . . . 294
7.2 Effectuer un transfert de données. . . . 295
Sélectionner la source du jeu de données. . . . 295
Description du jeu de données. . . . 295
Création d'une copie de jeu de données. . . . 306
Transfert de données. . . . 306
Transfert de manipulation de données. . . . 307
Sélection des variables. . . . 308
7.3 Obtenir des statistiques descriptives pour un jeu de données. . . . 308
Chargement du jeu de données d'entrée. . . . 308
Paramètres avancés du modèle. . . . 309
Analyse des statistiques. . . . 310
Tuiles. . . . 311
7.4 Répertorier des valeurs distinctes dans un jeu de données. . . . 315
Description de vos données. . . . 315
Définition des données d'entrée. . . . 316
Sortie. . . . 316
Paramètres. . . . 317

8 Définition des préférences. . . . 318
8.1 Présentation des menus. . . . 318
8.2 Préférences. . . . 319
Définition des préférences d'application. . . . 319
Paramétrage des options générales. . . . 319
Paramétrage des options de stockage. . . . 322
Définition de la délégation à APL pour l'entraînement des modèles. . . . 323
Configuration de la modélisation native Spark. . . . 324

Définition des options du référentiel de métadonnées. 325
Modification du contenu de la bibliothèque de variables. 325
Définition et modification d'une structure de variable. 327
Paramétrage des options de graphique. 330
Personnalisation de la feuille de style des rapports. 331
Paramétrage des options de géolocalisation. 334
8.3 À propos de. 334
8.4 Journal d'erreurs. 335
8.5 Contacter le Support. 335

1 Nouveautés de l'aide Automated Analytics

Liens dirigeant vers les informations concernant les nouvelles fonctionnalités et les modifications de la documentation pour l'aide en ligne Automated Analytics.

SAP Predictive Analytics 3.3

Nouveautés	Lien vers plus d'informations
Échantillonnage dans Data Manager	Échantillonnage [page 69]

1.1 Historique du document

Version du produit	Nouveautés
SAP Predictive Analytics 3.2	Processus de modélisation disponibles [page 135] SAP HANA en tant que source de données [page 84] Éditeur d'enregistrements analytiques > Accueil > Invites [page 25]
SAP Predictive Analytics 3.1	Des liens contenant plus d'informations sont ajoutés. Voir Où trouver des informations complémentaires ? [page 10] Voir Présentation de SAP Predictive Analytics [page 15]
SAP Predictive Analytics 3.0	Des liens sont ajoutés aux aides en ligne Predictive Factory et Expert Analytics. Voir Présentation de SAP Predictive Analytics [page 15]
	Les vues d'attribut SAP HANA sont prises en charge en tant que source de données. Voir SAP HANA en tant que source de données [page 84]
	Les modèles Predictive Factory peuvent être exposés dans Automated Analytics. Voir Ouverture ou suppression d'un modèle [page 79]

Version du produit	Nouveautés
	La description du graphique de prévisions expliquant pourquoi les premières valeurs du signal ne s'affichent pas est mise à jour. Voir Affichage des prévisions [page 196]
	Une note à propos du cluster supplémentaire contenant des enregistrements non affectés lors de la génération d'expressions SQL est ajoutée. Voir Synthèse des clusters [page 188]
	Le code CCL peut désormais être généré pour la classification, la régression et la mise en cluster des modèles. Voir Paramètres standard [page 218]
	La définition des groupes de variables est mise à jour. Voir Définitions [page 91]
SAP Predictive Analytics 2.5	Pour les modèles utilisant HIVE en tant que source de données, les calculs d'entraînement des modèles par défaut sont délégués au moteur Spark. Voir Configuration de la modélisation native Spark [page 324]
	Des informations ont été ajoutées sur l'option Événement cible requis pour les modèles Social. Voir Définitions de graphe [page 236]
SAP Predictive Analytics 2.4	Pour les modèles utilisant SAP HANA en tant que source de données, les calculs d'entraînement des modèles par défaut sont délégués à la bibliothèque SAP Automated Predictive (APL) depuis SAP HANA (lorsque cela est possible). Voir Définition de la délégation à APL pour l'entraînement des modèles [page 323]
	La définition de la valeur anormale pour un modèle de série temporelle a été mise à jour. Voir Affichage des prévisions [page 196]
SAP Predictive Analytics 2.3	La prise en charge des vues SAP HANA en tant que source de données de modélisation dans Automated Analytics (Classification/Régression, Mise en cluster, Séries temporelles, Règles d'association) est élargie pour inclure Social et Recommandation. Voir SAP HANA en tant que source de données [page 84]

Version du produit	Nouveautés
	Utilisation d'une vue d'analyse ou de calcul SAP HANA en tant que source de données pour votre modèle. Voir Sélection d'une vue SAP HANA [page 88]
	Utilisation d'une table SAP HANA ou une vue SQL en tant que source de données pour votre modèle. Voir Sélection d'une table SAP HANA ou d'une vue SQL en tant que source de données [page 89]

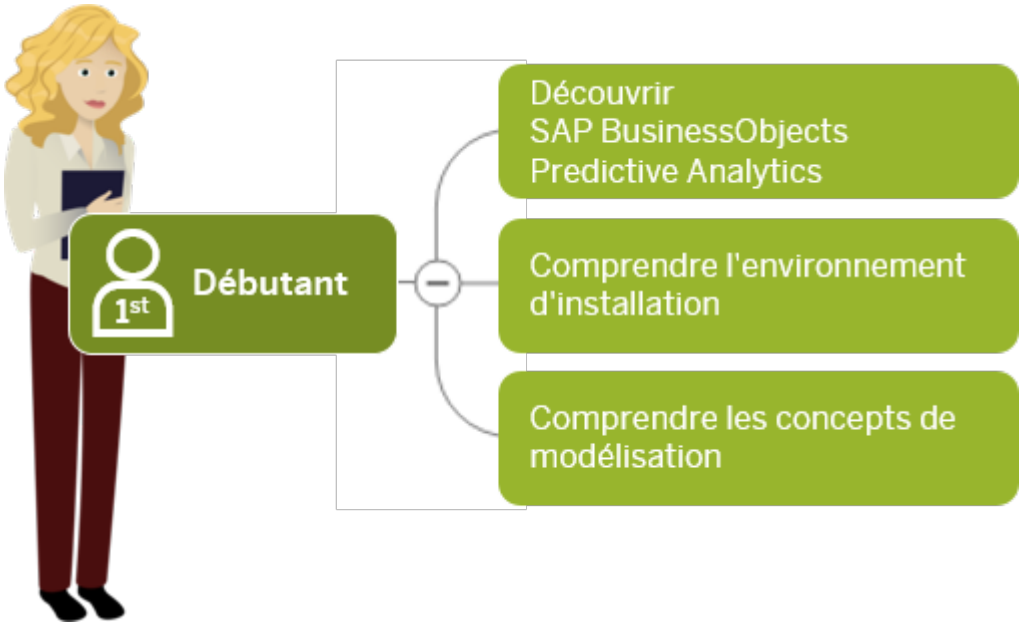
2 Où trouver des informations complémentaires ?

Cette aide en ligne vous apporte des informations relatives à l'utilisation de Automated Analytics.

Si vous recherchez des informations plus approfondies, ou si vous ne savez pas par où commencer, cliquez sur le profil qui vous correspond.

- Ce que vous devez savoir avant de commencer à utiliser SAP Predictive Analytics [page 11]
- Ce que vous devez savoir en tant qu'administrateur [page 12]
- Ce que vous devez savoir en tant qu'analyste [page 13]

2.1 Ce que vous devez savoir avant de commencer à utiliser SAP Predictive Analytics



- #unique_19/unique_19_Connect_42_subsection-im1 [page 11]
- #unique_19/unique_19_Connect_42_subsection-im2 [page 11]
- #unique_19/unique_19_Connect_42_subsection-im3 [page 11]

Placez le curseur de la souris au-dessus de chaque action pour obtenir une description. Cliquez sur l'action pour obtenir plus d'informations.

Découvrir SAP Predictive Analytics

Découvrez SAP Predictive Analytics, sa technologie et son architecture.

Présentation de SAP Predictive Analytics

Comprendre l'environnement d'installation

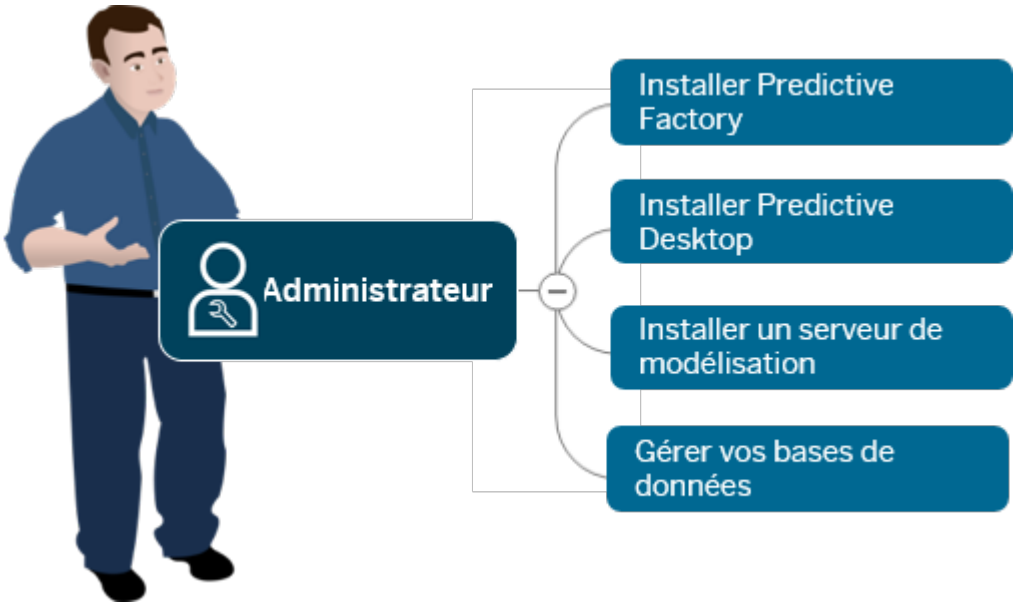
Utilisez l'un des quatre scénarios de déploiement pour vous guider tout au long de l'installation de SAP Predictive Analytics.

Comprendre les concepts de modélisation

Si vous débutez dans la modélisation avec SAP Predictive Analytics, lisez la section suivante pour découvrir tout ce que vous devez savoir pour commencer :

Concepts de base

2.2 Ce que vous devez savoir en tant qu'administrateur



- #unique_20/unique_20_Connect_42_subsection-im1 [page 12]
- #unique_20/unique_20_Connect_42_subsection-im2 [page 12]
- #unique_20/unique_20_Connect_42_subsection-im3 [page 12]
- #unique_20/unique_20_Connect_42_subsection-im4 [page 12]

Placez le curseur de la souris au-dessus de chaque action pour obtenir une description. Cliquez sur l'action pour obtenir plus d'informations.

Installer Predictive Factory

Si vous souhaitez installer et configurer Predictive Factory, lisez le guide d'installation et de configuration.

Installer Desktop

Lisez le guide d'installation Desktop.

Installer un serveur de modélisation

Apprenez à installer des serveurs Automated Analytics :

- Sous Linux
- Sous Windows

Apprenez à installer le client

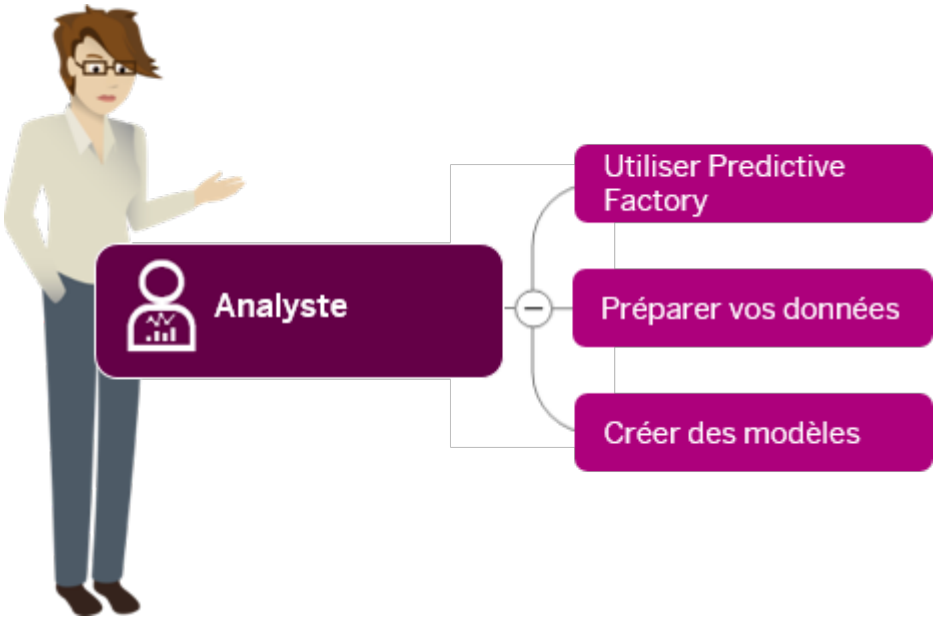
Gérer les connexions à vos bases de données

- Connecter votre base de données sous Linux
- Connecter votre base de données sous Windows

- Optimiser vos connexions aux bases de données

Pour obtenir plus d'informations sur les connexions aux bases de données, consultez la section Connectivity du SAP Help Portal.

2.3 Ce que vous devez savoir en tant qu'analyste



- #unique_21/unique_21_Connect_42_subsection-im1 [page 13]
- #unique_21/unique_21_Connect_42_subsection-im2 [page 13]
- #unique_21/unique_21_Connect_42_subsection-im3 [page 14]

Placez le curseur de la souris au-dessus de chaque action pour obtenir une description. Cliquez sur l'action pour obtenir plus d'informations.

Utiliser Predictive Factory

Lisez l'aide en ligne de Predictive Factory.

Préparer vos données

Découvrez la méthodologie qui se cache derrière Data Manager et apprenez à préparer vos données avant de créer vos modèles à l'aide de scénarios pratiques.

- Présentation des manipulations de données
- Scénario de manipulation de données
- Scénario d'agrégat sur périodes de temps
- Scénarios d'analyse des séquences

- Scénario d'analyse de texte

Créer des modèles

- Scénario de classification et de régression
- Scénario de segmentation et de mise en cluster
- Scénarios de séries temporelles
- Guide de l'utilisateur de l'analyse des réseaux sociaux
- Scénario de règle d'association
- Guide de l'utilisateur de Expert Analytics

3 Panneau de démarrage

3.1 Présentation de SAP Predictive Analytics

Predictive Factory utilise la valeur commerciale des modèles prédictifs, assurant le lien avec les bases de données productives qui fournissent les données aux systèmes opérationnels. Son interface conviviale permet aux utilisateurs (des analystes aux scientifiques des données) de créer, d'opérationaliser et de surveiller les modèles prédictifs, dans un workflow sécurisé et productif, via :

- Une gestion tout-en-un des modèles prédictifs associés à un contexte professionnel : création de modèles ; importation des modèles existants créés dans Automated Analytics ou Expert Analytics ; création directe de modèles de prévision de séries temporelles ; gestion des versions de modèles ; suivi des modèles pendant tout leur cycle de vie.
- Une planification précise des tâches principales d'industrialisation : application des modèles sur les nouvelles données, réentrainement des modèles pour assurer leur niveau de performance et leur précision, détection des écarts de modèles.

Veuillez vous référer à l'Aide en ligne de Predictive Factory

SAP Predictive Analytics est une solution d'exploration de données et de modélisation prédictive qui permet de découvrir des informations et relations cachées au sein de vos données et de créer des modèles prédictifs à partir desquels vous pouvez faire des prédictions relatives à des événements futurs.

Automated Analytics inclut les modules suivants :

- Data Manager sert à faciliter la préparation des données qui devront être utilisées dans le projet d'analyse.
- Modeler permet aux analystes de créer des modèles en suivant un workflow homogène et maniable ; par exemple une classification, une régression, une segmentation, une série temporelle ou des règles d'association. Les modèles peuvent être exportés sous différents formats afin que vous puissiez facilement les appliquer dans votre environnement de production.
- Social extrait et utilise des informations relationnelles structurelles stockées dans différents types de jeux de données, améliorant ainsi les fonctionnalités de décision et de prévision des modèles. Les données peuvent être représentées sous forme de graphiques montrant le lien entre les différentes données. Les workflows dédiés vous aident à créer des analyses des chemins fréquents et des colocations en fonction de données géoréférencées.
- Recommendation génère des recommandations de produits pour vos clients selon une analyse de liens fournie par Social.

Veuillez vous référer à Automated Analytics l'aide en ligne.

Expert Analytics permet :

- de produire une analyse approfondie des données à l'aide de diverses techniques de visualisation, telles que les diagrammes matriciels à nuage de points, les coordonnées parallèles, les diagrammes de cluster et les arbres de décision ;
- de réaliser différentes analyses et de créer des modèles sur les données, notamment la prévision de série temporelle, la détection de valeurs hors norme, l'analyse de tendance, l'analyse de classification, l'analyse de segmentation et l'analyse d'affinité ;

- d'utiliser une série d'algorithmes prédictifs, le langage d'analyse statistique en open source R et les fonctionnalités d'extraction de données en mémoire pour la gestion efficace de l'analyse de grands volumes de données.

Veuillez vous référer à l'aide en ligne de Expert Analytics

3.2 À propos de Data Manager

Data Manager vous permet de préparer vos données au traitement dans Automated Analytics. Il comprend les fonctionnalités suivantes :

- *Créer ou modifier des données analytiques*
  Cette option vous permet de créer à partir de vos données des objets analytiques tels que des populations estampillées, des entités, des indicateurs de performance clés (KPI), des enregistrements analytiques et de les stocker afin de pouvoir les réutiliser.
  Pour en savoir plus sur les données analytiques, consultez le guide Scénario de manipulation de données.
- *Créer une manipulation de données*
  Cette option vous permet de créer des manipulations de données que vous pourrez utiliser comme source de données dans toutes les fonctionnalités d'Automated Analytics.
  Pour en savoir plus sur les manipulations de données, consultez le guide Présentation des manipulations de données.
- *Charger une manipulation de données existante*
  Cette option vous permet de recharger une manipulation de données afin de la modifier.
- *Agrégats sur périodes de temps*
  L'outil d'agrégats sur périodes de temps proposé par Data Manager est un outil de prétraitement qui permet de réunir deux tables de bases de données relationnelles en une seule. Par exemple, il peut rassembler un fichier de clients statique et un fichier de transactions vertical en un fichier horizontal. Les agrégats peuvent être créés pour divers intervalles (jours, semaines, mois, trimestres, etc.). Les données rassemblées sont alors traitées à l'aide d'un modèle de classification ou de segmentation. Les agrégats ne sont pas stockés dans une nouvelle table, et aucun schéma de table n'est créé.

  > **i** Remarque
  >
  > Les variables date doivent être formatées comme suit : `YYYY/MM/DD`

  Pour en savoir plus sur les agrégats sur périodes de temps, consultez le guide Scénario d'agrégat sur périodes de temps. Veuillez noter que des exemples de données peuvent être téléchargés à partir de la page Help Portal pour réaliser le scénario proposé dans ce guide.
- *Séquences d'événements*
  La séquence d'événements est un outil de prétraitement qui regroupe des événements en une série de transitions. Par exemple, l'ensemble des clics d'un client sur un site web peut être transformé en une série de données pour chaque session. Chaque colonne représente une transition spécifique d'une page à une autre. Comme pour l'outil d'agrégats sur périodes de temps, ces nouvelles colonnes de données peuvent être ajoutées aux données client préexistantes et être utilisées par Automated Analytics Modeler.
  Pour en savoir plus sur l'analyse des séquences, consultez le guide Scénarios d'analyse des séquences.
  Veuillez noter que des exemples de données peuvent être téléchargés à partir de la page Help Portal pour réaliser le scénario proposé dans ce guide.

- *Codage textuel*
  Le codage textuel est un outil de prétraitement qui analyse et quantifie les variables textuelles. En association avec les capacités de régression d'Automated Analytics Modeler, cet outil identifie les mots importants et les affiche selon leur degré de pertinence par rapport à votre problématique métier. Cette fonctionnalité multilingue peut être personnalisée et enrichie pour s'adapter à différentes langues ou domaines d'application.
  Pour en savoir plus sur les agrégats sur périodes de temps, consultez le guide Scénario d'analyse de texte.

3.3 À propos de Automated Analytics

Automated Analytics offre les fonctionnalités suivantes :

- *Modeler* permet de créer des modèles tels que la classification, la régression, l'organisation en clusters, la série chronologique, les règles d'association, etc.
- *Social* améliore les capacités de décision et de prédiction de vos modèles à l'aide de l'extraction et de l'utilisation d'informations relationnelles structurelles implicites stockées dans différents types de jeux de données. Cela vous permet également d'utiliser les informations géographiques de vos données définies en exécutant des colocations ou des analyses des chemins fréquents.
- *Recommandation* génère des recommandations de produits pour vos clients en fonction de l'analyse du réseau social fournie par l'application.

Modeler

Modeler propose les fonctionnalités suivantes :

- *Créer un modèle de classification/régression*
  Cette fonctionnalité utilise un algorithme de régression qui trouvera le meilleur compromis entre l'ajustement et la robustesse du modèle. La robustesse est définie comme étant la cohérence d'un modèle de formation pour travailler aussi bien sur de nouvelles données.
  Pour en savoir plus sur la classification et la régression, consultez le guide Scénarios de classification, de régression, de segmentation et de mise en cluster.
- *Créer un nouveau modèle d'organisation en clusters*
  Cette fonctionnalité utilise un algorithme d'organisation en clusters qui trouvera un nombre de clusters pour un utilisateur spécifié dans un mode non surveillé ou dans un mode surveillé dans lequel une variable cible est sélectionnée.
  Pour en savoir plus sur la segmentation et la mise en cluster, consultez le guide Scénarios de classification, de régression, de segmentation et de mise en cluster.
- *Créer une analyse de série chronologique*
  Cette fonctionnalité vous permet de construire des modèles prédictifs à partir d'une série chronologique représentant des données. Grâce à ces modèles, vous pouvez identifier et comprendre le phénomène représenté par votre série chronologique et prévoir l'évolution sur le court et le moyen terme.
  Pour en savoir plus sur les séries temporelles, consultez le guide Scénarios de séries temporelles.
- *Créer des règles d'association*

Cette fonctionnalité utilise un algorithme générant des règles d'association à partir d'une table de référence et d'une table de transaction. Les règles d'association révèlent les relations entre les produits ou les services et suggèrent immédiatement les actions appropriées.
Pour en savoir plus sur les règles d'association, consultez le guide Présentation des règles d'association.

- *Charger un modèle*
  Cette option vous permet de recharger un modèle généré.

Social

- *Créer une analyse des réseaux sociaux*
  Social est un outil de prétraitement qui extrait et utilise les informations relationnelles structurelles implicites stockées dans différents types de jeux de données, améliorant ainsi les capacités de prise de décision et de prédiction des modèles. Les données peuvent être représentées sous forme de graphiques montrant le lien entre les différentes données.
- *Créer une analyse de colocation*
  Utilisez la fonction de géolocalisation pour construire un modèle vous permettant de visualiser les personnes qui se trouvaient au même emplacement géographique au même moment.
- *Créer une analyse de chemin fréquent*
  Construisez un modèle qui extrait les chemins/séquences les plus communs à partir d'événements géoréférencés.
- *Charger un modèle d'analyse de réseau social*
  Cette option vous permet de recharger un modèle généré.

Pour en savoir plus sur l'analyse des réseaux sociaux, consultez le guide Présentation des réseaux sociaux.

Recommandation

- *Créer un recommandeur*
  Recommendation vous permet de générer des recommandations pour vos clients en fonction de l'analyse du réseau social fournie par Automated Analytics Social.
- *Charger un conseiller*
  Cette option vous permet de recharger un conseiller généré précédemment.

3.4 À propos d'Expert Analytics

Grâce à Expert Analytics, vous pouvez réaliser différentes analyses sur les données, notamment la prévision de séries temporelles, la détection de valeurs hors norme, l'analyse de tendance, l'analyse de classification, l'analyse de segmentation et l'analyse d'affinité. Cette application permet d'analyser les données à l'aide de diverses techniques de visualisation, telles que les diagrammes matriciels à nuage de points, les coordonnées parallèles, les diagrammes de cluster et les arbres de décision.

Expert Analytics propose une gamme d'algorithmes prédictifs, prend en charge l'utilisation du langage d'analyse statistique en open source R et propose des fonctionnalités d'extraction de données en mémoire pour un traitement d'analyse efficace de grands volumes de données.

3.5 À propos de Toolkit

Les fonctionnalités proposées dans *Toolkit* vous permettent :

- d'explorer et de supprimer des jeux de données (*Ouvrir le visualiseur de données*) ;
- de transférer une source de données vers un autre emplacement ou dans un autre format (*Effectuer un transfert de données*) ;
- d'exporter la liste de valeurs distinctes d'une variable depuis votre jeu de données. Cette fonctionnalité vous permet entre autres de générer une liste de valeurs indexées unique à partir d'un jeu de données existant et de l'utiliser comme table de référence pour la journalisation des événements ou le codage des séquences (*Répertorier les valeurs distinctes dans un jeu de données*) ;
- de générer des statistiques des variables pour un jeu de données, telles que les statistiques croisées ou les courbes de profit lorsqu'une variable dans le jeu de données est utilisée pour évaluer la variable cible. Ce panneau vous permet de sélectionner le jeu de données que vous souhaitez analyser (*Obtenir des statistiques descriptives pour un jeu de données*) ;
- d'utiliser le moteur de codage pour regrouper, selon une cible, les catégories et les vignettes de variables lors de l'utilisation de données géographiques (*Calculer les groupes optimisés pour une cible*).

Pour en savoir plus sur Toolkit, consultez le guide Présentation de Toolkit.

4 Data Manager

4.1 Connexion à un référentiel d'objets analytiques

Ce panneau vous permet de sélectionner le référentiel de métadonnées auquel vous voulez vous connecter. Un référentiel de métadonnées est toujours une connexion à une base de données.

Définition

Un référentiel de métadonnées est un référentiel persistant d'objets analytiques de données.

Sélection d'un référentiel d'objets analytiques

- Pour sélectionner le référentiel de métadonnées
  1. Cliquez sur le bouton *Parcourir* correspondant au champ *Source de base de dionnées*. La fenêtre *Sélection d'une source de données* s'affiche.
  2. Sélectionnez la base de données contenant le référentiel de métadonnées à utiliser.
  3. Si la base de données à laquelle vous voulez accéder est protégée par mot de passe, entrez le nom d'utilisateur dans le champ **Utilisateur** et le mot de passe dans le champ **Mot de passe**.

     > **i** Remarque
     >
     > Si votre identifiant/mot de passe n'est pas valide, un message d'erreur s'affiche lorsque vous tentez de sélectionner une table.

  4. Cliquez sur *OK* pour valider votre sélection.
  5. Cliquez sur *Suivant* pour afficher le panneau *Gestion des données analytiques*.

4.2 Définition de la nouvelle manipulation de données

Ce panneau vous permet de sélectionner la table de base de données qui vous servira à créer un jeu de données d'analyse.

- Pour sélectionner une table
  1. Cliquez sur le bouton *Parcourir* correspondant au champ *Source de données*. La fenêtre *Sélection d'une source de données* s'affiche.

2. Sélectionnez la base de données contenant la table à partir de laquelle vous souhaitez créer un jeu de données d'analyse.
3. Si la base de données à laquelle vous voulez accéder est protégée par mot de passe, entrez le nom d'utilisateur dans le champ *Nom* et le mot de passe dans le champ *Mot de passe*.

> **i** Remarque
>
> Si vous entrez un nom et/ou mot de passe erronés, un message d'erreur s'affiche lorsque vous essayez de sélectionner une table dans cette base de données.

4. Cliquez sur le bouton *OK* pour valider votre sélection.
5. Cliquez sur le bouton *Parcourir* correspondant au champ *Table*. La fenêtre *Sélection d'une source de données* s'affiche.
6. Sélectionnez dans la liste la table que vous souhaitez utiliser dans le nouveau jeu de données d'analyse.
7. Cliquez sur le bouton *OK* pour valider votre sélection.
8. Dans le champ *Alias de la nouvelle table*, saisissez le nom par lequel vous voulez vous référer à la nouvelle table. Par défaut, l'alias est constitué à partir du nom de la table d'origine.
9. Cliquez sur le bouton *Suivant* pour afficher le panneau *Éditeur de jeux de données*.

4.3 Gestion des données analytiques

Ce panneau vous permet de :

- visualiser le contenu de votre entrepôt d'objets analytiques ;
- créer des objets ;
- supprimer les objets existants.

- Définitions des objets [page 21]
- Actions sur les objets [page 23]
- Options d'affichage [page 23]

4.3.1 Définitions des objets

Entité

Une entité est l'objet visé par une tâche d'analyse : cela peut être un client, un produit, un magasin, et est généralement identifié par un identifiant unique qui peut être utilisé dans tout le référentiel d'objets analytiques.

Enregistrement analytique

Un enregistrement analytique est une vue logique de tous les attributs correspondant à une entité. Il peut être décomposé en domaines qui regroupent des attributs reliés entre eux : ainsi, en CRM, un enregistrement analytique peut contenir un domaine démographique et un domaine comportemental.

Une population estampillée est une liste de paires `<identifiers; time stamps>` : qui représentent des instantanés des entités à un moment donné. Une entité peut figurer plusieurs fois dans une même population estampillée mais associée à des moments différents.

Table d'analyse temporelle

- *Population estampillée filtrée*, sur laquelle vous pouvez appliquer des filtres.
- *Population estampillée par produit cartésien*, obtenue en croisant une entité et une table de dates, c'est-à-dire en associant chaque ligne de l'entité à chaque date de la table de date.
- *Composition de populations estampillées*, obtenue en utilisant les opérateurs ensemblistes (union, intersection et différence) sur des populations estampillées existantes.

Jeu de données d'analyse temporelle

La méthodologie de gestion des données analytiques proposée par Data Manager utilise un type spécifique de jeu de données d'analyse, appelé *jeu de données d'analyse temporelle*. Il s'agit du produit d'une population estampillée par un enregistrement analytique. Le résultat de cette opération peut être considéré comme une table virtuelle contenant les valeurs des attributs associées à des identifiants en fonction de l'heure. Autrement dit, Un jeu de données d'analyse contient des "photos" (ou instantanés) d'une liste donnée d'entités prises à un certain moment, qui peut être différent pour chaque entité et une même entité peut être associée à plusieurs "photos".

Indicateur de performance

Les indicateurs de performance sont des outils de définition et de mesure de la progression qui permettent aux entreprises d'atteindre leurs objectifs. Ils sont convenus avec l'entreprise et peuvent être mesurés pour refléter les facteurs de réussite. Les indicateurs de performance choisis doivent refléter les buts de l'entreprise, ils doivent être primordiaux à son succès et doivent être mesurables.

Il existe deux manières de créer un indicateur de performance :

- Soit *à partir de la table calendaire*, c'est-à-dire d'une table contenant une date pour chaque jour avec des informations de calcul supplémentaires sur la date, comme le jour de la semaine, le jour de l'année, la semaine de l'année, etc. Cette option vous permet de créer un indicateur de performance à partir de zéro en créant des champs (agrégats, fonctions ...), en appliquant des filtres, en fusionnant des tables, etc.
- Soit *à partir d'une série temporelle*, c'est-à-dire d'une table contenant déjà un signal auquel vous pourrez ajouter des informations ou appliquer des filtres.

L'option avancée *Croiser une table avec des dates* vous permet de croiser une table de référence avec une table contenant une colonne de date pour obtenir une ligne par référence pour chaque date trouvée. Elle peut être

utilisée lors de la création d'indicateurs de performance avancés afin de pouvoir utiliser la colonne de date comme colonne de jointure lors des agrégations.

4.3.2 Actions sur les objets

- Pour créer un objet analytique :
  1. Cliquer sur le bouton **+** qui se trouve en haut à gauche du panneau. Une liste déroulante s'affiche alors contenant les objets qu'il est possible de créer.
  2. Cliquez sur le type d'objet que vous voulez créer. Un éditeur spécialisé s'affiche alors.
- Pour modifier un objet existant :
  1. Sélectionnez un objet dans la liste. L'état du bouton *Modifier* vous indique si l'objet peut être modifié ou non :

	L'objet peut être modifié.
	L'objet est utilisé par un autre objet et ne peut être modifié.

  2. Cliquez sur le bouton *Modifier* ou cliquez deux fois sur l'objet. Un éditeur spécifique s'affiche.

- Pour supprimer un objet :
  1. Sélectionnez un objet dans la liste. Si l'objet peut être supprimé le bouton **-** est dégrisé. Un objet utilisé par un autre objet ne peut pas être supprimé.
  2. Cliquez sur le bouton **-** ou sur la touche `Suppr`.
- Pour visualiser les données d'un objet analytique :
  1. Sélectionnez un objet analytique dans la liste des objets existant.
  2. Cliquez sur le bouton *Voir les données*. Si l'objet sélectionné nécessite une date de référence, une fenêtre s'ouvre vous permettant de paramétrer la date.

4.3.3 Options d'affichage

Changer de vue

Le bouton **Changer de vue** vous permet de choisir le mode d'affichage de la liste des objets analytiques.

L'option...		Affiche la liste comme...
	Dans une table	une table qui peut être triée par colonne.

L'option...		Affiche la liste comme...
	Dans une arborescence	une arborescence qui liste des objets par type.

Options de filtrage

Le bouton *Options de filtrage* vous permet de sélectionner ce qui doit apparaître dans la liste. Vous pouvez filtrer la liste par :

- *Types de connecteurs* : décochez les cases correspondant aux types de connecteurs que vous ne souhaitez pas voir apparaître dans la liste.
- *Nom* : saisissez le nom ou une partie du nom correspondant aux connecteurs que vous voulez voir apparaître dans la liste. Vous pouvez utiliser le caractère étoile (*) pour remplacer une partie de mot ou un point d'interrogation (?) pour remplacer un caractère unique.
- *Version* : choisissez si vous souhaitez afficher toutes les versions des objets, seulement la dernière version ou bien des versions spécifiques.

Cliquez sur le bouton *Fermer* pour enregistrer les options de filtrage et mettre à jour la liste en conséquence.

4.4 Éditeur d'enregistrements analytiques

Ce panneau vous permet de définir les paramètres d'un jeu de données d'analyse ou d'un indicateur de performance.

Le ruban situé en haut du panneau est divisé en trois parties : l'onglet *Accueil* permettant de modifier les champs, de créer des jointures et des filtres, l'onglet *Vues*, qui vous permet d'afficher les données ainsi que le SQL généré, et pour finir, l'onglet *Édition*, grâce auquel vous pouvez sélectionner et modifier les domaines.

- Accueil [page 25]
- Vues [page 31]
- Édition [page 38]

4.4.1 Accueil

L'onglet *Accueil* vous permet de modifier les champs en définissant des agrégats par exemple pour fusionner certains champs, créer des filtres ou définir une invite.

Éditeur d'expressions

L'éditeur d'expressions vous permet de créer des champs, un par un ou plusieurs à la fois, et de modifier les conditions de filtrage comme vous le feriez avec un calculateur.

Pour créer votre expression, plusieurs éléments sont à votre disposition :

- *Fonctions* : elles vous permettent de créer des expressions complexes avec un ou plusieurs champs.
- *Variables* et leurs valeurs :
  - les *champs* de votre base de données ;
  - les *ensembles de champs* que vous avez définis lors de votre manipulation de données ;
  - les *invites* que vous avez définies lors de votre manipulation de données ;
  - les *catégories* définies pour les champs ou variables existants.
- *Associations de champs*.
- *Messages*.

Définir un agrégat

Ce panneau vous permet d'ajouter des agrégats du même type que ceux créés automatiquement par la fonctionnalité de journalisation des événements. Ainsi, vous pouvez utiliser cette fonctionnalité pour générer automatiquement de nombreux agrégats différents, puis utiliser un modèle de classification ou de régression pour identifier les plus importants par rapport à votre question métier. Vous pouvez ensuite créer un jeu de données ne contenant que les agrégats les plus intéressants.

Paramètres d'agrégation

- Pour sélectionner la table des événements

Dans la section *Sélection de la table des événements* :

1. Sélectionnez dans la liste *Table* la table contenant les événements liés à votre table de référence.
2. Dans la liste *Colonne de date*, sélectionnez le champ indiquant l'ordre des événements.

- Pour spécifier les clés de jointure

Dans la section *Clés de jointure* :

1. Sélectionnez dans la liste *Clé de la table de référence* le champ correspondant à la clé de jointure.
2. Sélectionnez dans la liste *Clé de la table des événements* le champ correspondant à la clé de jointure.

- Pour spécifier l'opération d'agrégation
  1. Dans la section *Spécification de l'opération d'agrégation* :
  2. Sélectionnez dans la liste *Fonction* le type de fonction à appliquer. Les fonctions suivantes sont proposées :

Fonctions	Description	Valeurs renvoyées
Count	calcule le nombre d'occurrences	nombre d'occurrences trouvées
Sum	calcule la somme	somme calculée
Average	calcule la moyenne	moyenne calculée
Min	identifie la valeur minimum	valeur minimum
Max	identifie la valeur maximum	valeur maximum
Exists	vérifie s'il existe au moins un événement correspondant à la référence actuelle	0 si aucun événement n'a été trouvé 1 si au moins un événement a été trouvé
NotExists	vérifie qu'aucun événement correspondant à la référence actuelle n'existe	0 si au moins un événement a été trouvé 1 si aucun événement n'a été trouvé
Premier	identifie la première occurrence Remarque : Nécessite la présence d'une colonne de date	valeur de la première occurrence dans le temps pour la référence actuelle
Dernier	identifie la dernière occurrence Remarque : Nécessite la présence d'une colonne de date	valeur de la dernière occurrence dans le temps pour la référence actuelle

- Dans la liste *Colonne cible*, sélectionnez la variable sur laquelle appliquer la fonction choisie. Lorsque vous sélectionnez l'opération *Compter*, l'option supplémentaire * permet de ne pas sélectionner une colonne spécifique.
  Vous pouvez sélectionner plusieurs colonnes cibles.

Paramètres de période

Vous pouvez créer des agrégats sur l'ensemble du jeu de données ou bien le filtrer par date. Les dates servant de filtres peuvent être des constantes, des variables ou des invites.

- Pour paramétrer les périodes
  1. Cochez la case *Définir les périodes* pour sélectionner les dates de début et de fin déterminant la période utilisée pour filtrer les événements.
  2. Choisissez si vous voulez créer une *période simple* ou des *périodes successives*.
  3. *Dans le cas d'une période simple* :
     1. Choisissez le type d'entrée à utiliser comme date de début. Trois types sont disponibles :

- le type *Champ* vous permet d'utiliser une date stockée dans la base de données et donc propre à chaque client (par exemple une date de premier achat ou de résiliation d'abonnement) ;
- le type *Constante* vous permet de sélectionner une date fixe (par exemple la date du lancement d'une campagne marketing) ;
- le type *Invite* permet de demander à l'utilisateur de renseigner l'information au moment de la génération du modèle.

2. Sélectionnez le format de stockage à utiliser (*Date avec horaire* ou *Date*). Si vous avez sélectionné le type *Champ*, seuls les champs qui correspondent au format de stockage sélectionné sont affichés dans la liste.
3. Sélectionnez la valeur à utiliser comme date de début.
   - Dans le cas d'un champ, sélectionnez le champ dans la liste affichée.
   - Dans le cas d'une constante, saisissez la valeur de la constante ou utilisez le bouton calendrier à droite du champ.
   - Dans le cas d'une invite, sélectionnez l'invite à utiliser ou cliquez sur le bouton + pour créer une invite.

   > **i Remarque**
   >
   > La date de début est incluse dans la période.
4. Répétez les étapes 2 à 4 pour la date de fin.

   > **i Remarque**
   >
   > La date de fin est exclue de la période.

Dans le cas de périodes successives :

- Définissez le nombre de périodes successives que vous voulez, leur durée et la date de début en utilisant les liens hypertextes (soulignés en bleu).

Filtres et pivots

L'option *Filtre* vous permet de filtrer vos données selon les valeurs d'une variable. Chaque filtre défini génère une variable.

L'option *Pivot* vous permet de créer une variable pour chaque valeur sélectionnée. La création d'un pivot correspond à créer un filtre sur une catégorie pour chaque catégorie sélectionnée. Le pivot est toujours appliqué à la table des événements.

Les deux options peuvent être complémentaires.

- Pour définir un filtre

Il existe deux types de filtres :

- *Filtrer la table d'événements* affiche la liste des filtres appliqués à la table des événements.
- *Filtrer la table de référence* affiche la liste des filtres appliqués à la table de référence.

1. Cliquez sur *Ajouter* (✚) pour définir une condition complexe à l'aide de l'éditeur d'expressions.

2. Sélectionnez la condition définie précédemment et cliquez sur *Modifier* ( ) pour la modifier.
3. Cliquez sur *Supprimer* ( ) pour supprimer la condition définie précédemment.

- Définir un pivot
  1. Dans la liste déroulante *Variables*, sélectionnez la variable sur laquelle vous voulez filtrer.
  2. Pour ajouter des catégories à la table, vous pouvez :
     - extraire automatiquement les catégories des variables en cliquant sur les jumelles ( ) situées à côté de la liste, puis sélectionner les valeurs en cochant la case *Sélection* correspondante ;
       ou
     - saisir une valeur dans le champ *Nouvelle catégorie* et cliquez sur le bouton + ;
     - charger une liste de catégories de fichiers en cliquant sur le bouton Ouvrir le fichier ( ) à droite de la liste de tables Catégories, puis sélectionner les valeurs à conserver ou à exclure en cochant la case *Sélection* correspondante.

     > **i Remarque**
     >
     > Le nombre de variables créées est indiqué en bas du panneau. Ce nombre augmente de façon exponentielle en cas de filtrage par pivot. Plus le nombre de variables est élevé, plus l'apprentissage est long.

  3. Enregistrez les catégories sélectionnées en cliquant sur le bouton Enregistrer ( ) à droite de la liste de tables Catégories.

     > **i Remarque**
     >
     > Comme indiqué précédemment, le bouton *Enregistrer* enregistre uniquement les catégories dont la case *Sélection* est cochée.

  4. Cochez la case *Créer aussi des agrégats non filtrés* pour créer une variable agrégée non filtrée.

Filtrage des événements

Vous pouvez créer des agrégats sur l'ensemble du jeu de données ou bien le filtrer par date. Comme dans la journalisation des événements, les dates servant de filtre peuvent être des constantes, des variables ou des invites.

- Pour filtrer des événements
  1. Cochez la case *Fenêtre de temps* pour sélectionner les dates de début et de fin de la période utilisée pour filtrer les événements.
  2. Choisissez le type d'entrée à utiliser comme date de début. Trois types sont disponibles :
     - le type *Champ* vous permet d'utiliser une date stockée dans la base de données et donc propre à chaque client (par exemple une date de premier achat ou de résiliation d'abonnement)
     - le type *Constante* vous permet de sélectionner une date fixe (par exemple la date du lancement d'une campagne marketing)
     - le type *Invite* permet de demander à l'utilisateur de renseigner l'information au moment de la génération du modèle.

3. Sélectionnez le format de stockage à utiliser (*Date avec horaire* ou *Date*). Si vous avez sélectionné le type *Champ*, seuls les champs qui correspondent au format de stockage sélectionné sont affichés dans la liste.
4. Sélectionnez la valeur à utiliser comme date de début.
   - Dans le cas d'un champ, sélectionnez le champ dans la liste affichée.
   - Dans le cas d'une constante, saisissez la valeur de la constante ou utilisez le bouton calendrier à droite du champ.
   - Dans le cas d'une invite, sélectionnez l'invite à utiliser ou cliquez sur le bouton **+** pour créer une invite.
5. Répétez les étapes 2 à 4 pour la date de fin.
6. Cliquez sur *Suivant* pour créer l'agrégat. Une fenêtre s'ouvre vous permettant de nommer le nouveau champ.
7. Dans le champ *Nom* saisissez le nom du champ à créer.
8. Cliquez sur *OK*.

Jointures

Cet onglet vous permet de créer une jointure entre la table source et une autre table de votre base de données, autrement dit d'ajouter les informations contenues dans une autre table lorsque le champ sélectionné dans la table source est identique au champ sélectionné dans la table cible.

- Pour définir une jointure
  - Table
    1. Sélectionnez la table à joindre dans la liste *Table cible*.
    2. Cliquez sur le bouton +.
    3. Sélectionnez le champ source dans la liste *Champ source*.
       Si la table sélectionnée contient des champs correspondant au champ source, ceux-ci sont affichés dans la liste *Champ cible*. Sinon, le message *Aucun champ disponible* est affiché.
    4. Sélectionnez le champ cible dans la liste *Champ cible*.
    5. Cliquez sur *OK* pour créer la jointure. Ce bouton est activé uniquement lorsque tous les éléments nécessaires à la jointure sont sélectionnés. Les jointures existantes sont affichées dans la partie supérieure du panneau.

Une fois qu'une jointure a été créé, tous les champs de la table cible sont ajoutés à la liste *Champ source*, ce qui vous permet de créer des jointures à partir de ces champs.

> **i** Remarque
>
> Si vous choisissez de créer une jointure sans utiliser de champ clé, un icône d'avertissement s'affiche dans la liste des jointures existantes.

- Champs

Lorsque vous fusionnez de nouveaux champs, vous pouvez en définir vous-même l'appellation en ajoutant un préfixe ou un suffixe aux alias.

- Dans le champ *Préfixe*, saisissez le préfixe de votre choix, suivi d'un trait de soulignement.
- Dans le champ *Suffixe*, saisissez le suffixe de votre choix, précédé d'un trait de soulignement.

Après avoir fusionné deux champs à partir de tables différentes, vous avez la possibilité de créer une clé composite en joignant deux autres champs de ces mêmes tables.

- Pour créer une clé composite
  1. Sélectionnez le champ source dans la liste *Champ source*.
  2. Sélectionnez le champ cible dans la liste *Champ cible*.
  3. Cliquez sur *Nouvelle fusion*.
     Un menu contextuel apparaît.
  4. Sélectionnez l'option *Ajouter une paire de clés*.
     Les champs source et cible de la jointure existante, qui apparaissent dans la partie supérieure du panneau, sont mis à jour.

Filtres

Cet onglet vous permet de créer et d'appliquer des filtres sur les données pour ne conserver que les enregistrements qui vous intéressent.

- Pour définir un filtre
  1. Dans la section *Ne conserver que les entrées qui rendent vrai pour*, choisissez si vous voulez conserver les enregistrements répondant à toutes les conditions ou seulement à au moins l'une d'elles.
  2. Cliquez sur *Nouvelle condition*. L'éditeur d'expressions s'ouvre, vous permettant ainsi de définir une nouvelle condition.
- Pour modifier une condition
  1. Dans la *Liste des conditions du filtre*, sélectionnez la condition que vous voulez modifier.
  2. Cliquez sur *Modifier*.
- Pour supprimer une condition de filtre
  1. Dans la *Liste des conditions du filtre*, sélectionnez la condition que vous voulez supprimer.
  2. Cliquez sur *Supprimer*.

Invites

Cet onglet liste les invites existantes et permet de définir de nouvelles invites qui serviront à demander à l'utilisateur d'entrer une valeur lors de l'utilisation du jeu de données d'analyse dans un composant d'Automated Analytics. Deux types d'invite peuvent servir aux manipulations de données : invites créées ou invites héritées Les **invites créées** sont définies directement dans Data Manager selon la description de la procédure suivante. Elles peuvent être modifiées dans Data Manager et servir lors de la définition de champs, de filtres et de fusions. Les **invites héritées** proviennent d'une vue SAP HANA. Elles ne peuvent être ni modifiées ou supprimées dans Data Manager, ni utilisées pour la définition de nouveaux champs, filtres ou fusions.

- Pour créer une invite

Il existe deux façons de créer une invite, soit dans l'onglet *Invite*, soit lors de la création d'un champ ou d'un filtre.

1. Dans l'*Éditeur d'invites*, saisissez le nom de l'invite dans le champ *Nom*. Ce nom permet de sélectionner l'invite sous forme de valeur lors de la création d'un champ ou d'un filtre.

2. Choisissez le type d'invite dans la liste *Type*.
3. Saisissez la valeur par défaut dans le champ *Valeur*. Lorsque l'utilisateur voit l'invite apparaître, cette valeur est proposée par défaut.
4. Saisissez la phrase qui demande à l'utilisateur de saisir une valeur dans le champ *Description*.
5. Cliquez sur le bouton *OK* pour créer l'invite.

4.4.2 Vues

Utilisez cet onglet pour afficher les données et leurs statistiques ou pour visualiser l'expression SQL et la documentation générée correspondant à la manipulation de données en cours. Si la manipulation de données contient des invites créées ou héritées, elles sont affichées et vous devez saisir les valeurs demandées.

Champs

Cet onglet affiche les champs existant et vous permet d'ajouter vous propres champs.

Les champs existants peuvent être modifiés grâce aux opérations suivantes :

- Modifier l'alias d'un champ.
- Définir la visibilité d'un champ, autrement dit indiquer si le champ sera visible dans le jeu de données d'analyse ou non. Les champs indiqués comme non visibles n'apparaissent pas dans l'onglet *Données et statistiques*, mais peuvent être utilisés dans une jointure ou un filtre par exemple.
- Ajouter une Description.
- Modifier les informations d'un champ (uniquement dans le cas d'un champ créé par l'utilisateur).
- Pour créer un champ calculé
  1. Cliquez sur le bouton *+ Nouveau*.
  2. Sélectionnez le type de champ que vous voulez créer.
  3. Définissez les paramètres de champ en fonction du type de champ que vous avez sélectionné. Le champ créé apparaît dans la liste avec les champs d'origine de la table source.

Données et statistiques

Cette option affiche les données extraites de la base de données quand cette manipulation de données est exécutée. Elle fournit aussi certaines statistiques relatives à ces données. Il vous permet de vérifier si les résultats correspondent à votre attente. Si la manipulation de données contient des invites, qu'elles soient créées dans Data Manager ou héritées d'une source SAP HANA, vous devez les renseigner pour afficher les données et statistiques.

Vous pouvez trier la table en cliquant sur l'en-tête de la colonne correspondante.

SQL généré

Cet onglet affiche la requête SQL correspondant au jeu de données d'analyse en cours de création.

Documentation

L'onglet *Documentation* vous permet d'accéder à une synthèse de votre manipulation de données. Il contient toutes les options sélectionnées pour votre manipulation de données comme Filtres, Jointure, Invites ou Expressions.

Cet écran présente :

- Aperçu graphique
- Champs visibles/non visibles
- Invite
- Expressions
- Filtres

L'aperçu graphique présente :

- la liste des tables incluses dans votre manipulation de données, et comment ces dernières ont été créées (avec des jointures et agrégats par exemple)
- les champs conservés pour votre manipulation de données, connus en tant que Champs visibles.

Champs visibles affiche dans un tableau les colonnes des champs visibles, etc.

Vous pouvez personnaliser cet aperçu dans les options de paramétrage.

- Paramètres des sections
- Paramètres des champs

Les paramètres de synthèse vous permettent de choisir le format de la synthèse, à la fois pour l'affichage et l'exportation . Le fichier généré peut être enregistré aux formats de fichier txt, .htm et .rtf.

Infos

Le bouton *Infos* vous permet d'afficher les informations techniques relatives à la base de données qui héberge vos données. Les informations suivantes sont disponibles :

- le nom de la base de données, son numéro de version et le numéro de version du driver
- la taille maximale autorisée pour les noms de tables et de colonnes
- le nombre maximal de colonnes autorisé dans les tables et les requêtes
- la taille maximale autorisée pour une requête
- le nombre maximal de tables autorisé dans une requête

Options

Options	Description	Valeurs/Défaut
Optimiser les expressions utilisées plusieurs fois	Dans une manipulation de données, deux stratégies peuvent être utilisées lorsqu'un champ calculé ou une expression est référencé plusieurs fois : • La *stratégie standard*, dans laquelle chaque référence est remplacée par la définition SQL complète du champ calculé en question, le calcul étant ainsi effectué plusieurs fois. • La *stratégie optimisée*, dans laquelle une colonne temporaire est calculée une seule fois en utilisant la définition SQL du champ. Ensuite chaque référence suivante à ce champ est remplacée par une référence à cette nouvelle colonne.	Automatique (par défaut) : la stratégie est sélectionnée automatiquement par le système. Par défaut, si un champ est référencé plus d'une fois, la stratégie optimisée est utilisée. Activé : la stratégie optimisée est utilisée. Désactivé : la stratégie standard est utilisée.
Nombre d'utilisations max. de l'expression avant optimisation	Lorsque l'option Optimiser les expressions utilisées plusieurs fois est activée, cette option vous permet de spécifier à partir de combien de références à un champ la stratégie optimisée doit être utilisée.	Par défaut ce seuil est de 2, ce qui signifie que la stratégie optimisée est utilisée dès qu'un champ est référencé plus d'une fois.
Générer des requêtes multiples	Cette option permet de forcer ou de désactiver la décomposition d'une manipulation de données en étapes intermédiaires (définies automatiquement par le système).	Automatique (par défaut) : selon la complexité relative de la manipulation de données, le système décide d'activer ou non la décomposition. Activé : la manipulation de données est décomposée en étapes intermédiaires, quelle que soit sa complexité. Désactivé : la manipulation de données n'est jamais décomposée en étapes intermédiaires, quelle que soit sa complexité. Cette valeur doit être sélectionnée avec précaution.
Générer les agrégats standard en tant que tables corrélées	Cette option permet de forcer ou de désactiver la génération d'agrégats standards (min, max, count, avg) à l'aide de tables corrélées plutôt que de sous-requêtes de type SELECT.	Automatique (par défaut) : le système sélectionne automatiquement le type de génération d'agrégats utilisé.

Options	Description	Valeurs/Défaut
Générer les agrégats 'First/Last' en tant que tables corrélées	Cette option vous permet de paramétrer l'utilisation de tables corrélées plutôt que de sous-requêtes de type SELECT lors de la génération des agrégats de type 'First/Last'.	Activé : la forme Tables corrélées est générée lorsque le système de gestion de base de données relationnelle le permet. Désactivé : la forme Sous-requête est générée lorsque le système de gestion de base de données relationnelle le permet. Certaines bases de données ignorent ces options lorsqu'elles ne prennent en charge qu'une seule des deux formes. Ainsi, MySQL prend uniquement en charge les sous-requêtes SELECT alors que Teradata ne peut traiter que les tables corrélées. Les formes SQL prises en charge par les différentes bases de données sont répertoriées dans la section Formes SQL des agrégats prises en charges par les bases de données.
Générer les agrégats 'Exists' en tant que tables corrélées	Cette option vous permet de paramétrer l'utilisation de tables corrélées plutôt que de sous-requêtes de type SELECT lors de la génération des agrégats de type 'Exists'.	
Générer les conditions de jointure après chaque requête	Par souci de clarté, la notation (T1,T2) est utilisée pour représenter l'opération de jointure externe gauche entre la table T1 et T2 dans certaines conditions. Dans une longue chaîne ou séquence de jointures de tables, par exemple ((T1, (T2,T4)), (T3,T4)), certaines conditions de jointure peuvent être générées soit directement après la jointure, soit à la fin de la requête sans que cela n'impacte l'ensemble des lignes obtenues comme résultat. Cependant dans certains cas, la position des conditions en fin de requête peut impacter négativement les performances du SQL généré. Cette option vous permet de forcer la position des conditions de jointure.	Automatique (par défaut) : le système sélectionne automatiquement la forme préférée en fonction des expériences effectuées sur la base de données concernée. Activé : les conditions de jointure sont générées immédiatement après l'opération de jointure. Désactivé : les conditions de jointure sont générées à la "fin" de la séquence de jointure.

Options	Description	Valeurs/Défaut
Matérialiser les populations estampillées	Dans un environnement opérationnel, les jeux de données d'analyse d'Automated Analytics génèrent du code SQL complexe que certains optimiseurs de SGBDR (systèmes de gestion de base de données relationnelle) peuvent parfois ne pas analyser correctement. Cela a pu être observé en particulier lorsque des populations estampillées filtrées sont concernées. Pour contourner cela, la matérialisation de la population estampillée aide l'optimiseur à prendre de meilleures décisions pour le plan d'exécution. Cette option vous permet de paramétrer ce comportement.	Automatique (par défaut) : le système utilise des règles spécifiques pour décider automatiquement d'appliquer ou non la matérialisation. Actuellement, toutes les populations estampillées filtrées sont matérialisées. Activé : toutes les populations estampillées sont matérialisées. Désactivé : les populations estampillées ne sont jamais matérialisées.
Grouper les agrégats avec pivots (par valeur de pivot)	Dans des bases de données telles que Teradata, Oracle, et DB2, les agrégats sont formulés en tant qu'expressions de tables corrélées ou dérivées. Chacune de ces tables rassemble les agrégats générés sur le même ensemble de lignes. L'ensemble de ligne est défini d'une part par les colonnes clés et d'autre part par la condition de filtrage. Le filtre comprend les spécifications de la fenêtre de temps (lorsqu'elle est spécifiée) ainsi qu'un prédicat supplémentaire. Cette option vous permet de générer ce prédicat afin qu'il soit pris en compte sans avoir d'impact négatif sur la condition de filtrage. Cela implique de générer les agrégats avec pivots dans des tables différentes.	Automatique (par défaut) : le système décide d'ajouter ou non le prédicat supplémentaire à la condition de filtrage en fonction du nombre d'agrégats associés. Le nombre seuil d'agrégats peut être défini grâce à l'option Seuil de groupement des filtres et pivots similaires décrite ci-dessous. Activé : le prédicat supplémentaire ne fait jamais partie de l'expression de filtre. Désactivé : le prédicat supplémentaire fait toujours partie de l'expression de filtre.
Seuil de groupement des filtres et pivots similaires	Lorsque l'option Grouper les agrégats avec pivots (par valeur de pivot) a la valeur Automatique, les agrégats ayant la même fenêtre de temps mais ayant des conditions de filtrage différentes sont regroupés si leur nombre dépasse un certain seuil. Cette option permet de paramétrer ce seuil.	Le seuil par défaut est 5.

Options	Description	Valeurs/Défaut
Nombre d'agrégats avec pivots par groupement	Lorsque l'option Grouper les agrégats avec pivots (par valeur de pivot) a la valeur Automatique, les agrégats ayant la même fenêtre de temps mais ayant des conditions de filtrage différentes sont regroupés. Ce paramètre vous permet de définir le nombre maximal d'agrégats pouvant constituer un groupe.	Le seuil par défaut est 7.
Pré-filtrage des agrégats	Les expressions de tables regroupant des agrégats avec pivots peuvent filtrer les lignes vues par ces agrégats. Ce pré-filtrage concerne souvent des colonnes non-indexées, ce qui conduit à des balayages complets. Lorsque toutes les colonnes utilisées dans les filtres d'agrégats ou dans les conditions de jointure sont identifiées comme étant indexées, activer le pré-filtrage peut avoir un impact positif sur les performances. Chaque agrégat traitant uniquement les lignes correspondantes, ce pré-filtrage peut être laissé désactivé.	Par défaut, cette option est désactivée.
Champs toujours visibles	Les objets de données analytiques, tels que les entités ou les populations estampillées, sont configurés pour n'exposer que certains champs. Cette option vous permet de spécifier si d'autres champs doivent être également visibles aux utilisateurs de la manipulation de données.	Par défaut les champs KxTarget et KxWeight sont visibles. Pour ajouter des champs supplémentaires, saisissez leurs noms séparés par un point-virgule.
Remplacer les champs constants par leurs valeurs	À des fins d'optimisation, lorsqu'une manipulation de données doit exposer un champ de valeur constante, la valeur réelle est propagée et les manipulations de données supérieures sont informées que le champ concerné est constant et que sa valeur doit donc être utilisée là où le nom du champ est référencé. Cela permet à certaines manipulations de données d'être exécutées 10 fois plus vite.	Cette option est activée par défaut. Pour la désactiver, décochez la case.

Options	Description	Valeurs/Défaut
Optimiser les expressions utilisant les champs d'une sous-manipulation de données	Dans une manipulation de données, les expressions référencées plusieurs fois sont factorisées par souci d'efficacité. Cependant, lorsque ces expressions utilisent certains champs issus d'une jointure ou d'une sous manipulation de données, le système peut ne pas les factoriser correctement. Cette option vous permet d'exclure ces expressions de la factorisation. Remarque : Cette option n'est utilisée que si les options Optimiser les expressions utilisées plusieurs fois et Nombre d'utilisations max. de l'expression avant optimisation sont activées.	Cette option est désactivée par défaut. Pour l'activer, cochez la case.
Ne pas exclure les clés hors-jointure des optimisations intermédiaires	Par souci d'optimisation, la factorisation d'expressions référencées plusieurs fois n'est effectuée que lorsque la table sur laquelle elle est calculée contient des colonnes clés. Dans les manipulations de données à plusieurs niveaux, pour éviter de violer les contraintes de non-nullité et d'unicité des clés physiques, le système applique des règles spécifiques supprimant les clés qui ne sont pas utilisées dans les tables intermédiaires. De ce fait, certaines expressions référencées plusieurs fois peuvent ne pas être factorisées. Cette option permet de forcer le système à ne pas exclure des tables intermédiaires les colonnes clés pouvant être utilisées pour l'optimisation.	Cette option est activée par défaut. Pour la désactiver, décochez la case.

Formes SQL des agrégats prises en charges

	Standard	First/Last	Exists
Access	Sous-requête uniquement	Sous-requête uniquement	Sous-requête uniquement

	Standard	First/Last	Exists
SQLServer	Tables corrélées par défaut	Tables corrélées par défaut	Sous-requête uniquement
Oracle	Tables corrélées par défaut	Tables corrélées uniquement	Sous-requête uniquement
Teradata	Tables corrélées uniquement	Tables corrélées uniquement	Tables corrélées uniquement
DB2	Sous-requête uniquement	Tables corrélées uniquement	Sous-requête uniquement
MySQL	Sous-requête uniquement	Sous-requête uniquement	Sous-requête uniquement
PostgreSQL	Tables corrélées par défaut	Tables corrélées uniquement	Tables corrélées uniquement
SybaseIQ	Tables corrélées uniquement	Tables corrélées uniquement	Tables corrélées uniquement
Vertica	Tables corrélées uniquement	Tables corrélées uniquement	Tables corrélées uniquement
Netezza	Tables corrélées uniquement	Tables corrélées par défaut	Tables corrélées par défaut

Légende :

- *Sous-requête uniquement* : dans cette base de données, les agrégats sont générés en utilisant uniquement des sous-requêtes.
- *Tables corrélées uniquement* : dans cette base de données, les agrégats sont générés en utilisant uniquement des tables corrélées.
- *Tables corrélées par défaut* : dans cette base de données, des tables corrélées sont générées par défaut mais la génération de sous-requêtes `SELECT` peut aussi être forcée lorsque cela est possible.

4.4.3 Édition

La section *Édition* permet de sélectionner et modifier des champs.

Sélection

Option	Description
Sélection avancée	Sélectionne les champs selon divers critères tels que leur alias, le type, l'emplacement de stockage ou la table dont ils proviennent.
Sélectionner tout	Sélectionne tous les champs de la manipulation de données.
Inverser la sélection	Inverse la sélection actuelle. Les champs sélectionnés sont désélectionnés et les champs non sélectionnés sont sélectionnés.

Modification du champ

Modification	Procédure
Renommer les champs	1. Sélectionnez les champs à renommer. 2. Cliquez sur *Renommer*. 3. Sélectionnez la modification à appliquer dans la liste déroulante. Les options suivantes sont disponibles : • *En majuscule* • *Tronquer* • *Ajouter un préfixe* • *Ajouter un suffixe* • *Remplacer la chaîne*
Définir la visibilité des champs	1. Sélectionnez les champs dont vous voulez modifier la visibilité. 2. Cliquez sur *Définir la visibilité*. 3. Sélectionnez l'option à utiliser dans la liste déroulante. Par défaut, seuls les champs visibles sont affichés dans la liste. Pour afficher les champs invisibles, décochez l'option *Afficher uniquement les champs visibles* située sous la liste des champs.
Définir le type de champ	1. Sélectionnez les champs dont vous voulez modifier le type. 2. Cliquez sur *Changer le type*. 3. Sélectionnez un nouveau type dans la liste déroulante. Les types suivants sont disponibles : • *nominal* • *ordinal* • *continu* • *textuel* Vérifiez que le type sélectionné correspond au contenu réel des champs que vous avez choisis.

4.5 Éditeur d'expressions

L'éditeur de formule vous permet de créer des champs (un par un ou en masse) et de modifier des conditions comme vous le feriez avec une calculatrice.

Fonctions d'agrégat de base

L'éditeur d'expressions prend en charge l'utilisation des fonctions d'agrégat de base, ce qui vous permet de regrouper vos données en colonnes pour simplifier et réduire le jeu de données.

Les fonctions d'agrégat de base ne s'affichent pas dans l'interface utilisateur de l'éditeur d'expressions par défaut. Pour accéder aux fonctions d'agrégat de base, vous pouvez :

- Ajouter l'option `<-BasicAggregation>` à la ligne de commande au démarrage de l'application de la manière suivante : `C:\Program Files\SAP Predictive Analytics\Desktop 3.1\Automated\EXE\Clients\KJWizardJNI\KJWizardJni.exe-BasicAggregation.`
- Ou ajouter la ligne suivante au fichier KJWizardJni.ini : `arg.1=-BasicAggregation`. Ce fichier se trouve sous le dossier : ...`\SAP Predictive Analytics\Desktop 3.1\Automated\EXE\Clients\KJWizardJNI\KJWizardJni.ini`.

Les fonctions d'agrégat suivantes sont prises en charges par SAP Predictive Analytics :

Fonctions d'agrégat de base	Formules
Somme	AggregateSum()
Min	AggregateMin()
Max	AggregateMax()
Nombre	AggregateCount()
Moy.	AggregateAvg()

Pour utiliser les fonctions d'agrégat de base dans l'éditeur d'expressions, vous pouvez saisir la formule ou sélectionner les fonctions d'agrégat de base correspondantes.

Exemple

Vous avez un jeu de données important qui contient beaucoup de données. Vous souhaitez agréger un indicateur, par exemple `<Quantité>` ou `<Montant>` par `<Temps>` et `<Produit>`, puis utiliser les résultats dans *Modeler* pour la *Prévision de séries temporelles* :

1. Ajoutez une nouvelle variable `<Quantité totale>` avec une formule de fonction d'agrégat de base : `AggregateSum(sales_orders.quantity_ordered).`
2. Cochez la case Maintenir la visibilité de chaque variable requise pour le regroupement : `<Date de début de mois>` et `<ID produit>`.

Création d'un champ

- Pour créer un champ :
  1. Entrez une formule dans la zone de texte qui se trouve dans la partie supérieure du panneau. Vous pouvez utiliser les fonctions d'aide à la saisie suivantes :
     - Cliquez deux fois sur le nom d'une fonction située dans le cadre *Fonctions* ou d'une variable (champ ou invite ou ensemble de champs prédéfini) située dans le cadre *Variables* pour l'insérer à

l'endroit du curseur. Dans le cas d'une fonction, un patron de la fonction avec des informations sur les paramètres à renseigner est inséré.

- Vous pouvez également insérer ces éléments à la position voulue en utilisant le glisser-déposer depuis l'un des arbres vers la zone de texte.
- Lors de la saisie d'un nom de variable (champ, invite ou ensemble), appuyez simultanément sur les touches `Ctrl` + `Espace` pour afficher une liste des noms de variable commençant par le texte saisi.
  La couleur de la pastille située au-dessus de la zone *Messages* vous informe de l'état de la formule.

	Si la pastille est...	la formule...
	rouge	contient une erreur, qui est signalée dans la zone des messages. Il n'est pas possible de la valider (le bouton Suivant est désactivé).
	jaune	peut être validée mais certaines incohérences sont possibles et sont signalées dans la zone des messages.
	verte	est valide.

Dans le cas d'une erreur ou d'un avertissement, la zone *Message* située dans la partie inférieure du panneau vous donne des détails pour vous aider à comprendre la raison de l'erreur ou de l'avertissement.

2. Pour valider définitivement la formule, cliquez sur le bouton *Suivant*.
3. Saisissez un nom pour le nouveau champ sous *Nom*.
4. Cliquez sur le bouton *OK*.

Fonctionnalités avancées

Ensembles de champs

Il est courant de vouloir créer plusieurs champs en appliquant un même calcul à différent champs existants. L'éditeur de formule vous permet de faire cela grâce à l'utilisation des ensembles de champs.

Par exemple, l'utilisation des ensembles de champs vous permet d'additionner un grand nombre de champs, ou de calculer leur maximum.

Ensembles de champs nommés

- Pour créer un ensemble de champs nommé :
  1. Dans la section *Variables*, cliquez deux fois sur l'option *Ensemble de champs*. Un sous-arbre s'affiche, listant les ensembles de champs existants.
  2. Cliquez deux fois sur l'option *Créer un ensemble de champs...* Une fenêtre s'affiche listant tous les champs disponibles dans votre table d'analyse.
  3. Dans le champ *Masque d'alias*, saisissez un masque permettant de filtrer les champs par leur nom. Un masque est constitué d'une partie commune au nom de tous les champs que vous voulez voir

s'afficher dans la liste et du caractère étoile (*) qui permet de compléter les parties de nom qui diffèrent selon les champs. L'étoile peut être utilisée en début, milieu ou fin de mot et autant de fois que nécessaire.

4. Décochez les champs que vous ne souhaitez pas conserver dans l'ensemble de champs.
5. Dans le champ *Nom*, saisissez le nom du nouvel ensemble de champs.
6. Cliquez sur le bouton OK. La fenêtre se ferme et le nouvel ensemble de champs apparaît dans la liste sous l'élément *Ensemble de champs*.

- Pour modifier un ensemble de champs nommé :
  1. Cliquez sur l'ensemble de champs que vous souhaitez modifier pour le sélectionner.
  2. Cliquez avec le bouton droit de la souris sur l'ensemble de champs sélectionné. Un menu contextuel s'affiche.
  3. Cliquez sur l'option Modifier. La fenêtre d'édition des ensembles de champs s'affiche.

> **i** Remarque
>
> Modifier le nom d'un ensemble de champs revient à le dupliquer.

- Pour supprimer un ensemble de champs nommé :
  1. Cliquez sur l'ensemble de champs que vous souhaitez modifier pour le sélectionner.
  2. Cliquez avec le bouton droit de la souris sur l'ensemble de champs sélectionné. Un menu contextuel s'affiche.
  3. Cliquez sur l'option *Supprimer*.

Ensembles de champ créés à la volée

Lorsque vous ne souhaitez pas appliquer un calcul sur les champs dont les noms ont une racine commune, vous pouvez créer un ensemble de champs à la volée dans le champ de texte de la formule.

Un ensemble de champ créé à la volée est défini par un **masque**, correspondant à une partie fixe commune aux noms de tous les champs souhaités et à un caractère générique représentant la partie du nom qui change pour chaque champ. Trois caractères génériques peuvent être utilisés pour définir les ensembles de champs : le signe arobase (@), le signe dièse (#) et le signe du dollar ($). Seul un caractère générique peut être utilisé pour chaque ensemble de champ, et un même caractère générique ne peut pas être utilisé deux fois dans la même formule.

- Pour créer un ensemble de champs à la volée :
  - Dans le champ de texte de la formule, saisissez le masque correspondant aux champs auxquels vous voulez ajouter un calcul, par exemple income_@, puis utilisez-le comme champ standard.
- Pour utiliser plusieurs ensembles de champs :
  1. Saisissez une formule à l'aide des ensembles de champs tel que défini précédemment. La zone *Messages* située dans la partie inférieure du panneau indique le nombre de champs allant être créés.
  2. Lorsque vous utilisez plusieurs ensembles de champs dans la même formule, vous devez sélectionner dans la liste déroulante *Association des champs* la manière dont les champs de ces ensembles seront associés.
     - *Associer par position* : les champs de chaque ensemble sont associés en fonction de leur position dans la table de base de données.
     - *Associer par valeur* : les champs de chaque ensemble sont associés en fonction de la valeur représentée par les caractères génériques utilisés pour définir les ensembles de champ.
     - *Effectuer le produit cartésien* : tous les champs d'un ensemble sont associés à tous les champs d'un autre ensemble.

Par exemple : prenons le cas d'une table de base de données contenant les champs suivants organisés comme suit :
income_january
income_april
income_february
income_march
expenses_march
expenses_january
expenses_april
La formule suivante `income_@ - expenses_#` utilise deux ensembles de champs, un regroupant tous les champs `income_` et le deuxième avec tous les champs `expenses_` :

income_@	expenses_#
income_january	expenses_march
income_april	expenses_january
income_february	expenses_april
income_march	

L'option *Associer par position* entraîne les calculs suivants :

Calcul	Position
income_january - expenses_march	1
income_april - expenses_january	2
income_february - expenses_april	3

Si l'option *Associer par valeur* est sélectionnée, l'application essayera d'associer la valeur représentée par `@` et la valeur représentée par `#`, entraînant les calculs suivants :

Calcul	Valeurs @ et #
income_january - expenses_january	january
income_april - expenses_april	april
income_march - expenses_march	march

L'option *Effectuer le produit cartésien* entraîne les calculs suivants :

Calcul
income_january - expenses_march
income_january - expenses_january
income_january - expenses_april
income_april- expenses_march
income_april - expenses_january
income_april - expenses_april
income_february - expenses_march
income_february - expenses_january
income_february - expenses_april
income_march - expenses_march
income_march - expenses_january
income_march - expenses_april

3. Cliquez sur *Suivant* pour valider la création des champs. Une boîte de dialogue vous permettant de nommer les nouveaux champs s'ouvre.
4. Saisissez une racine commune à tous les champs venant d'être créés.
5. Si la formule utilise uniquement les ensembles de champ nommés ou si vous n'avez pas besoin de contrôler davantage l'appellation, passez à l'étape suivante. Si vous avez utilisé un ou plusieurs caractères génériques, vous pouvez les utiliser dans l'alias pour créer un modèle d'appellation. Les caractères génériques sont remplacés par les parties de nom correspondant.
6. Cliquez sur *OK* pour créer les champs.

Utilisation d'un ensemble de champs en tant que liste de variables pour une fonction n-aire.

Un ensemble de champs peut être utilisé pour définir une liste des arguments pour des fonctions n-aire, c'est-à-dire des fonctions qui comportent un nombre indéfini d'arguments. Lorsqu'un ensemble de champs est utilisé en tant qu'argument dans une fonction n-aire, et dans ce cas uniquement, vous devez encadrer le nom de l'ensemble avec des accolades **{ }** pour qu'il soit interprété comme une liste d'arguments.

Prenons par exemple le cas d'une table de base de données sont les champs income_april, income_may, income_june contiennent le chiffre d'affaires mensuel. Si vous souhaitez créer un champ contenant le chiffre d'affaire mensuel le plus élevé du trimestre, vous devez utiliser la formule suivante `greatestN({income_@})` qui utilise l'ensemble de champs `income_@` en tant que liste de variables, ce qui équivaut à la formule `greatestN(income_april ,income_may ,income_june)`. Toutefois, lorsque vous utilisez l'ensemble de champs tel quel dans la formule `greatestN(income_@)`, trois champs sont créés par les trois formules suivantes : `greatestN(income_april)`, `greatestN(income_may)` et `greatestN(income_june)`, qui ne peut être utilisé.

- Pour utiliser un ensemble de champs en tant que liste de variable :
  1. Créez un ensemble de champs (nommé ou à la volée) tel que décrit précédemment.

2. Dans le champ de texte de la formule, saisissez la formule contenant une fonction n-aire.
3. Saisissez le nom de l'ensemble de champs entre accolades `{ }` pour l'utiliser en tant qu'argument pour la fonction. Par exemple, la formule `greatestN({incomes})` créera un champ dont la valeur est la valeur la plus élevée des champs contenus dans l'ensemble de champs appelé incomes.

4.6 Définition d'un agrégat

Ce panneau vous permet d'ajouter des agrégats du même type que ceux créés automatiquement par la fonctionnalité de journalisation des événements. Ainsi, vous pouvez utiliser cette fonctionnalité pour générer automatiquement de nombreux agrégats différents, puis utiliser un modèle de classification ou de régression pour identifier les plus importants par rapport à votre question métier. Vous pouvez ensuite créer un jeu de données ne contenant que les agrégats les plus intéressants.

- Paramètres d'agrégation [page 45]
- Paramètres de période [page 46]
- Filtres et pivots [page 47]

4.6.1 Paramètres d'agrégation

- Pour sélectionner la table des événements

Dans la section *Sélection de la table des événements* :

1. Sélectionnez dans la liste *Table* la table contenant les événements liés à votre table de référence.
2. Sélectionnez dans la liste *Colonne de date* le champ indiquant l'ordre des événements.

- Pour spécifier les clés de jointure

Dans la section *Clés de jointure* :

1. Sélectionnez dans la liste *Clé de la table de référence* le champ correspondant à la clé de jointure.
2. Sélectionnez dans la liste *Clé de la table des événements* le champ correspondant à la clé de jointure.

- Pour spécifier l'opération d'agrégation
  1. Dans la section *Spécification de l'opération d'agrégation* :
  2. Sélectionnez dans la liste *Fonction* le type de fonction à appliquer. Les fonctions suivantes sont proposées :

Fonctions	Description	Valeurs renvoyées
Count	calcule le nombre d'occurrences	nombre d'occurrences trouvées
Sum	calcule la somme	somme calculée
Average	calcule la moyenne	moyenne calculée
Min	identifie la valeur minimum	valeur minimum

Fonctions	Description	Valeurs renvoyées
Max	identifie la valeur maximum	valeur maximum
Exists	vérifie s'il existe au moins un événement correspondant à la référence actuelle	0 si aucun événement n'a été trouvé 1 si au moins un événement a été trouvé
NotExists	vérifie qu'aucun événement correspondant à la référence actuelle n'existe	0 si au moins un événement a été trouvé 1 si aucun événement n'a été trouvé
Premier	identifie la première occurrence Remarque : Nécessite la présence d'une colonne de date	valeur de la première occurrence dans le temps pour la référence actuelle
Dernier	identifie la dernière occurrence Remarque : Nécessite la présence d'une colonne de date	valeur de la dernière occurrence dans le temps pour la référence actuelle

3. Sélectionnez dans la liste *Colonne Cible*, la variable sur laquelle appliquer la fonction choisie.
   Lorsque vous sélectionnez l'opération *Compter*, une option supplémentaire * vous permet de ne pas sélectionner une colonne donnée.
   Vous pouvez sélectionner plusieurs colonnes cibles.

4.6.2 Paramètres de période

Vous pouvez créer des agrégats sur l'ensemble du jeu de données ou bien le filtrer par date. Les dates servant de filtres peuvent être des constantes, des variables ou des invites.

- Pour paramétrer les périodes
  1. Cochez la case *Définir les périodes* pour sélectionner les dates de début et de fin déterminant la fenêtre de temps utilisée pour filtrer les événements.
  2. Choisissez si vous voulez créer une *période simple* ou des *périodes successives*.
  3. *Dans le cas d'une période simple* :
     1. Choisissez le type d'entrée à utiliser comme date de début. Trois types sont disponibles :
        - le type *Champ* vous permet d'utiliser une date stockée dans la base de données et donc propre à chaque client (par exemple une date de premier achat ou de résiliation d'abonnement)
        - le type *Constante* vous permet de sélectionner une date fixe (par exemple la date du lancement d'une campagne marketing)
        - le type *Invite* permet de demander à l'utilisateur de renseigner l'information au moment de la génération du modèle.
     2. Sélectionnez le format de stockage à utiliser (*Date avec horaire* ou *Date*). Si vous avez sélectionné le type *Champ*, seuls les champs qui correspondent au format de stockage sélectionné sont affichés dans la liste.

3. Sélectionnez la valeur à utiliser comme date de début.
   - Dans le cas d'un champ, sélectionnez le champ dans la liste affichée.
   - Dans le cas d'une constante, saisissez la valeur de la constante ou utilisez le bouton calendrier à droite du champ.
   - Dans le cas d'une invite, sélectionnez l'invite à utiliser ou cliquez sur le bouton **+** pour créer une invite.

     > **i** Remarque
     >
     > La date de début est incluse dans la période.
4. Répétez les étapes **2** à **4** pour la date de fin.

   > **i** Remarque
   >
   > La date de fin est exclue de la période.

Dans le cas de périodes successives :

- Définissez le nombre de périodes successives que vous voulez, leur durée et la date de début en utilisant les liens hypertextes (soulignés en bleu).

4.6.3 Filtres et pivots

L'option *Filtre* vous permet de filtrer vos données selon les valeurs d'une variable. Chaque filtre défini génère une variable.

L'option *Pivot* vous permet de créer une variable pour chaque valeur sélectionnée. La création d'un pivot correspond à créer un filtre sur une catégorie pour toutes les catégories sélectionnées. Le pivot est toujours appliqué à la table des événements.

Les deux options peuvent être complémentaires.

- Pour définir un filtre

Il existe deux types de filtres :

- Filtrer la table d'événements affiche la liste des filtres appliqués à la table des événements.
- Filtrer la table de référence affiche la liste des filtres appliqués à la table de référence.

1. Cliquez sur le bouton *Ajouter* ( ) pour définir une condition complexe à l'aide de l'éditeur d'expressions.
2. Sélectionnez la condition définie précédemment et cliquez sur le bouton *Modifier* ( ) pour la modifier.
3. Cliquez sur le bouton *Supprimer* ( ) pour supprimer la condition définie précédemment.

- Définir un pivot

1. Dans la liste déroulante *Variables*, sélectionnez la variable sur laquelle vous voulez filtrer.
2. Pour ajouter des catégories à la table, vous pouvez :
   - extraire automatiquement les catégories des variables en cliquant sur les jumelles ( ) situées à côté de la liste, puis sélectionner les valeurs en cochant la case *Sélection* correspondante ;
     ou
   - saisir une valeur dans le champ *Nouvelle catégorie* et cliquez sur le bouton **+** ;

- charger une liste de catégories de fichiers en cliquant sur le bouton Ouvrir le fichier ( ) à droite de la liste de tables Catégories, puis sélectionner les valeurs à conserver ou à exclure en cochant la case *Sélection* correspondante.

  > **i Remarque**
  >
  > Le nombre de variables créées est indiqué en bas du panneau. Ce nombre augmente de façon exponentielle en cas de filtrage par pivot. Plus le nombre de variables est élevé, plus l'apprentissage est long.

3. Enregistrez les catégories sélectionnées en cliquant sur le bouton Enregistrer ( ) à droite de la liste de tables Catégories.

   > **i Remarque**
   >
   > Comme indiqué précédemment, le bouton Enregistrer enregistre uniquement les catégories dont la case Sélection est cochée.

4. Cochez la case *Créer aussi des agrégats non filtrés* pour créer une variable agrégée non filtrée.

Filtrage des événements

Vous pouvez créer des agrégats sur l'ensemble du jeu de données ou bien le filtrer par date. Comme dans la journalisation des événements, les dates servant de filtre peuvent être des constantes, des variables ou des invites.

- Pour filtrer des événements
  1. Cochez la case *Fenêtre de temps* pour sélectionner les dates de début et de fin de la période utilisée pour filtrer les événements.
  2. Choisissez le type d'entrée à utiliser comme date de début. Trois types sont disponibles :
     - le type *Champ* vous permet d'utiliser une date stockée dans la base de données et donc propre à chaque client (par exemple une date de premier achat ou de résiliation d'abonnement)
     - le type *Constante* vous permet de sélectionner une date fixe (par exemple la date du lancement d'une campagne marketing)
     - le type *Invite* permet de demander à l'utilisateur de renseigner l'information au moment de la génération du modèle.
  3. Sélectionnez le format de stockage à utiliser (*Date avec horaire* ou *Date*). Si vous avez sélectionné le type *Champ*, seuls les champs qui correspondent au format de stockage sélectionné sont affichés dans la liste.
  4. Sélectionnez la valeur à utiliser comme date de début.
     - Dans le cas d'un champ, sélectionnez le champ dans la liste affichée.
     - Dans le cas d'une constante, saisissez la valeur de la constante ou utilisez le bouton calendrier à droite du champ.
     - Dans le cas d'une invite, sélectionnez l'invite à utiliser ou cliquez sur le bouton **+** pour créer une invite.
  5. Répétez les étapes 2 à 4 pour la date de fin.
  6. Cliquez sur le bouton *Suivant* pour créer l'agrégat. Une fenêtre s'ouvre vous permettant de nommer le nouveau champ.

7. Dans le champ *Nom* saisissez le nom du champ à créer.
8. Cliquez sur le bouton *OK*.

4.7 Nouvelle condition

Une condition vous permet de définir au cas par cas la valeur d'un champ à partir des valeurs d'un autre.

- Pour créer une condition
  1. Dans la section *Définir un nouveau cas*, cliquez sur le bouton + situé à droite du champ *Si*. L'*éditeur de condition* s'affiche.
  2. Sélectionnez un opérateur pour la nouvelle condition dans la liste *Opérateur booléen*.
  3. Définissez les paramètres de condition.
  4. Cliquez sur le bouton *OK*. La condition apparaît dans le champ *Si*.
  5. Définissez quelle valeur prendra le nouveau champ lorsque la condition que vous venez de définir sera vraie.
  6. Cliquez sur le bouton *Ajouter un cas* et ajoutez le nouveau cas à la liste.
  7. Répétez les étapes 1 à 6 pour chaque cas à créer dans cette condition. L'ordre dans lequel les cas sont listés est important car la première condition vraie va déterminer la valeur du champ calculé. Utilisez les boutons et à gauche pour déplacer un cas dans la liste.
     Vous pouvez supprimer un cas en le sélectionnant puis en cliquant sur le bouton *Supprimer le cas*.
  8. Dans la section *Sinon*, définissez le paramètre à utiliser si aucun des cas définis n'est vrai.
  9. Cliquez sur le bouton *Suivant*. La fenêtre *Entrez le nom du champ calculé* s'affiche.
  10. Saisissez le nom du nouveau champ dans le champ *Nom*.
      Remarque : Si le nom saisi existe déjà, un message s'affiche en-dessous du champ Nom et le bouton *OK* est désactivé.
  11. Cliquez sur le bouton *OK*. La liste *Champs* est affichée avec le nouveau champ.

4.8 Ajout d'une nouvelle table de recherche

L'utilisateur précise des cas, chacun étant constitué d'une liste de valeurs discrètes correspondant à une étiquette. L'exemple classique est une table de recherche utilisée en tant que dictionnaire pour transformer des identifiants en chaînes de caractères, ou pour regrouper des valeurs représentant des distinctions précises en un nombre plus petit de classes générales.

- Pour créer une table de recherche
  1. Dans la liste *Champs*, sélectionnez celui qui contient les valeurs d'entrée du dictionnaire.
  2. Sélectionnez le type de résultats que vous voulez créer dans la liste *Format de sortie*.
  3. Dans le champ *Valeur*, saisissez la valeur que vous souhaitez ajouter comme entrée du dictionnaire.
  4. Cliquez sur le bouton + pour ajouter cette valeur à la liste *Valeurs possibles*. Si vous voulez que plusieurs valeurs d'entrée correspondent à la même valeur de sortie, répétez les étapes 3 et 4 pour chaque valeur.

5. Dans le champ *Alors*, saisissez la valeur de sortie correspondant aux valeurs d'entrée de la liste.
6. Cliquez sur le bouton *Ajouter* pour créer la nouvelle entrée. Le nouveau cas est affiché dans la liste.
7. Répétez les étapes 3 à 6 pour chaque entrée de la table de recherche.
8. Dans le champ *Sinon*, saisissez la valeur de sortie qui sera utilisée par défaut pour toutes les valeurs d'entrée qui n'ont pas été spécifiées dans la liste des cas.
9. Cliquez sur le bouton *Suivant*. La fenêtre *Entrez le nom du champ calculé* s'affiche.
10. Saisissez le nom du nouveau champ dans le champ *Nom*.

> **i** Remarque
>
> Si le nom saisi existe déjà, un message s'affiche et le bouton *OK* est désactivé.

11. Cliquez sur le bouton *OK*. La liste *Champs* est affichée avec le nouveau champ.

4.9 Nouvelle normalisation

La normalisation est une primitive de manipulation des données standard qui est utilisée en PMML (Predictive Model Markup Language), langage de spécification du data mining défini par le Data Mining Group (DMG). La normalisation est fréquemment appliquée aux variables numériques avant l'extraction des données. Il s'agit d'une transformation linéaire par morceau dont la variable résultante varie généralement entre 0 et 1. Elle peut être utilisée pour des transformations de rangs dans lesquelles le résultat représente la proportion approximative (percentile) de valeurs inférieures à la valeur d'entrée. Une autre possibilité consiste à convertir un champ en fonction de l'écart-type d'une valeur par rapport à la moyenne du champ. La normalisation comprend également l'indication de la valeur à utiliser quand une valeur d'entrée numérique est inconnue ou à l'extérieur du jeu de données d'entraînement.

- Pour créer une normalisation
  1. Sélectionnez un champ dans la liste déroulante *Choisissez un champ à normaliser*. Seuls les champs de type entier ou réel sont affichés.
  2. Pour chaque point à ajouter à la normalisation :
     1. saisissez sa valeur dans le champ *Entrée* de la section *Ajoutez des point à la normalisation*,
     2. Saisissez la valeur de sortie correspondante dans le champ *Sortie*.
     3. Cliquez sur le bouton + pour ajouter le point défini à la normalisation. Le point s'affiche dans la liste.
        Si deux points au moins ont été définis, le graphique correspondant s'affiche dans la section *Aperçu* en bas du panneau.
  3. Dans la liste *Valeurs minimales* de la section *Définir le comportement aux limites*, sélectionnez le comportement à appliquer aux valeurs inférieures au point le plus bas défini. Les valeurs disponibles sont détaillées dans le tableau ci-dessous.

Comportement	les valeurs extrêmes correspondent à ...	Courbe correspondante
Valeur constante	la valeur de la borne	IMAGE ICI

Comportement	les valeurs extrêmes correspondent à ...	Courbe correspondante
Pente	la suite vers l'infini de la droite passant par la dernière borne	IMAGE ICI
Défini par l'utilisateur	une valeur définie par l'utilisateur. Pour paramétrer une valeur, saisissez-la dans le champ de texte et cliquez sur le bouton d'actualisation.	IMAGE ICI
Valeur nulle	la valeur nulle. Ce qui signifie que les valeurs extrêmes n'apparaissent pas sur le graphique.	IMAGE ICI

4. Cliquez sur le bouton *Suivant*. La fenêtre *Entrez le nom du champ calculé* s'affiche.
5. Saisissez le nom du nouveau champ dans le champ *Nom*.

> **i** Remarque
>
> Si le nom saisi existe déjà, un message s'affiche et le bouton *OK* est désactivé.

6. Cliquez sur le bouton *OK*. La liste *Champs* est affichée avec le nouveau champ.

4.10 Nouvelles fonctions

Un très large éventail de fonctions peuvent intéresser les utilisateurs. Cependant, la plupart de ces fonctions sont construites à partir d'une ou plusieurs des fonctions suivantes :

- Opérateurs mathématiques : tels que +, *, -, /, %
- Opérateurs logiques : et, ou
- Manipulations de chaînes de caractères : comparaison, extraction, trim, trim à gauche, trim à droite, mise en majuscule, mise en minuscule.
- Fonctions mathématiques : log, exp
- Manipulations de données : année, mois, jour, heure, minute, seconde, addition de dates, différence de dates
  Remarque : La fonction DayOfTheWeek qui retourne le jour de la semaine, renvoie un entier qui varie de 1 à 7, où 1 représente dimanche, 2 représente lundi, et 7 représente samedi. Ce comportement est le comportement standard des principales bases de données relationnelles (telles que DB2, ORACLE, SQLServer, ACCESS).

La liste de base des manipulations de données est fournie par SQL.

- Pour créer une fonction
  1. Dans la liste *Fonctions*, sélectionnez celle que vous souhaitez utiliser pour créer le nouveau champ calculé. Les paramètres à remplir pour la fonction choisie sont affichés dans la deuxième partie du panneau.
  2. Définissez les paramètres de la nouvelle fonction.

3. Cliquez sur le bouton *Suivant*. La fenêtre *Entrez le nom du champ calculé* s'affiche.
4. Saisissez le nom du nouveau champ dans le champ *Nom*.

> **i** Remarque
>
> Si le nom saisi existe déjà, un message s'affiche et le bouton *OK* est désactivé.

5. Cliquez sur le bouton *OK*. La liste *Champs* est affichée avec le nouveau champ.

4.11 Nouvelles expressions SQL

Un champ contenant une expression libre vous permet d'utiliser des requêtes SQL prédéfinies.

- Pour définir une nouvelle expression libre
  1. Saisissez une expression SQL valide dans le champ *Expression SQL*. Si l'expression saisie n'est pas correcte, un message d'erreur s'affiche lorsque vous cliquez sur l'onglet *Voir les données*.
  2. Sélectionnez dans la liste *Type du résultat* le type de résultat désiré en sortie.
  3. Cliquez sur le bouton *Suivant*. La fenêtre *Entrez le nom du champ calculé* s'affiche.
  4. Saisissez le nom du nouveau champ dans le champ *Nom*.

  > **i** Remarque
  >
  > Si le nom saisi existe déjà, un message s'affiche et le bouton *OK* est désactivé.

  5. Cliquez sur le bouton *OK*. La liste *Champs* est affichée avec le nouveau champ.

4.12 Modification de jeu de données d'analyse temporelle

Ce panneau permet de créer un jeu de données d'analyse à partir d'un enregistrement analytique et d'une population estampillée, ou de modifier un jeu de données d'analyse existante.

Définition

Un jeu de données d'analyse est le produit d'une population estampillée par un enregistrement analytique. Le résultat de cette opération peut être considéré comme une table virtuelle contenant les valeurs des attributs associées à des identifiants en fonction de l'horodatage. Autrement dit, Un jeu de données d'analyse contient des "photos" (ou instantanés) d'une liste donnée d'entités prises à un certain moment, qui peut être différent pour chaque entité et une même entité peut être associée à plusieurs "photos".

Création ou modification d’un jeu de données d’analyse

- Pour créer ou modifier un jeu de données d’analyse
  1. Dans le champ *Nom* saisissez le nom que vous voulez attribuer au nouveau jeu de données d’analyse.
  2. Dans le champ *Description* saisissez une description du nouveau jeu de données d’analyse.
  3. Dans la liste déroulante *Enregistrement analytique*, choisissez l'enregistrement analytique à utiliser pour créer le jeu de données d’analyse.
  4. Dans la liste déroulante *Population estampillée*, choisissez la population estampillée à utiliser pour créer le jeu de données d’analyse.
  5. Cliquez sur le bouton *Suivant* pour valider la création ou les modifications du jeu de données d'analyse.
  6. Si un jeu de données d'analyse portant le même nom existe déjà, une boite de dialogue vous demande de préciser le comportement à adopter :
     - Cliquez sur *Oui* pour enregistrer le jeu de données d'analyse en tant que nouvelle version de celui qui existe.
     - Cliquer sur *Non* si vous voulez écraser le jeu de données d’analyse existant.

4.13 Modification d'un enregistrement analytique

Ce panneau vous permet de créer ou de modifier un enregistrement analytique.

Définition

Un enregistrement analytique est une vue logique de tous les attributs correspondant à une entité. Il peut être décomposé en domaines qui regroupent des attributs reliés entre eux : ainsi, en CRM, un enregistrement analytique peut contenir un domaine démographique et un domaine comportemental.

Création ou modification

- Pour créer ou modifier un enregistrement analytique
  1. Dans le champ *Nom*, saisissez le nom que vous voulez attribuer à votre enregistrement analytique.
  2. Dans le champ *Description* saisissez une description.
  3. Dans la liste déroulante *Basé sur l'entité* sélectionnez l'entité à partir de laquelle sera créé l'enregistrement analytique.
  4. Cliquez sur le bouton *Modifier les attributs* pour ajouter à l'enregistrement analytique des champs calculés ou issus d'autres tables.
  5. Facultatif : vous pouvez associer les attributs de l'enregistrement analytique à des domaines (voir la section Domaines ci-dessous). Par défaut, les attributs sont associés à un domaine nommé d'après la table dont il provient.

6. Cliquez sur le bouton *Suivant* pour valider la création ou la modification de l'enregistrement analytique.
7. Si un enregistrement analytique portant le même nom existe déjà, une boite de dialogue vous demande de préciser le comportement à adopter :
    - Cliquez sur *Oui* pour enregistrer l'enregistrement analytique avec un nouveau numéro de version.
    - Cliquer sur *Non* si vous voulez écraser l'enregistrement analytique existant.

Domaines

Définition

Un domaine regroupe des attributs du même type, tels que des informations démographiques ou comportementales. Le domaine comportemental est généralement obtenu grâce à des agrégats par entité sur des tables d'événements.

Un domaine est une unité sémantique utilisée uniquement pour les rapports d'affaires et les interprétations de modèles.

Modification des domaines

L'*éditeur d'enregistrements analytiques* vous permet de créer des domaines, de supprimer des domaines existants, de modifier leurs noms ou de changer la couleur qui leur est associée.

- Pour modifier les domaines
    - Dans le champ *Domaines* situé dans l'onglet Édition du ruban, cliquez sur le bouton *Modifier domaines*. Le panneau *Modifier domaines* s'ouvre en listant les domaines existants. Par défaut, un domaine est créé pour chaque table utilisée par l'enregistrement analytique.
- Pour créer un domaine
    1. Cliquez sur *Modifier domaines*.
    2. Cliquez sur le bouton + situé dans le coin supérieur droit de la fenêtre. Une ligne supplémentaire s'affiche dans le tableau contenant le nouveau domaine. Par défaut, celui-ci est nommé `<Domain>` suivi d'un numéro.
    3. Cliquez deux fois dans la cellule *Nom* correspondant au domaine créé pour le modifier, puis saisissez le nom du nouveau domaine.
- Pour changer la couleur d'un domaine
    1. Cliquez sur *Modifier domaines*.
    2. Cliquez sur la cellule *Couleur* correspond au domaine sélectionné. Une fenêtre de sélection de couleur s'affiche.
    3. Sélectionnez la couleur de votre choix dans l'onglet *Échantillons* ou utilisez les onglets *TSV* ou *RVB* si vous connaissez les valeurs exactes à saisir.
    4. Cliquez sur le bouton *OK* pour valider votre sélection.
- Pour supprimer un domaine
    1. Cliquez sur *Modifier domaines*.
    2. Sélectionnez le domaine à supprimer.
    3. Cliquez sur le bouton -.
- Pour associer des attributs à un domaine
    1. Dans le panneau *Modifier l'enregistrement analytique*, cliquez sur le bouton *Modifier les attributs...*. Le panneau *Enregistrement analytique* s'affiche.

2. Sélectionnez les attributs dans la table.
3. Dans le champ *Domaines* situé dans l'onglet *Édition* du ruban, cliquez sur le rectangle correspondant au domaine auquel vous voulez associer les attributs sélectionnés. Les attributs prennent la couleur du domaine choisi.

- Pour visualiser les domaines par couleur
  - Cochez la case *Montrer les domaines* pour faire apparaître les domaines par couleur.

4.14 Modification d'une entité

Ce panneau vous permet de créer une entité ou d'en modifier une.

Définition

Une entité est l'objet visé par une tâche d'analyse : cela peut être un client, un produit, un magasin, et est généralement identifié par un identifiant unique qui peut être utilisé dans tout le référentiel d'objets analytiques.

Création ou modification

- Pour créer une entité
  1. Dans le champ *Nom*, saisissez le nom que vous voulez attribuer à votre entité.
  2. Dans le champ *Description*, saisissez une description.
  3. Pour sélectionner la table à partir de laquelle vous voulez construire votre entité, cliquez sur le bouton *Parcourir* et sélectionnez une table dans la boîte de dialogue. Cliquez sur le bouton *OK* pour valider votre sélection.

Pour afficher les données de la table sélectionnée, cliquez sur le bouton (*Aperçu*).

Une confirmation peut vous être demandée avant l'exécution de la requête si vous utilisez une base de données Teradata(1), Oracle(2) ou SQLServer 2005. Cela est dû au *mode Explain* qui estime la charge de tous les requêtes SQL avant de les exécuter. Pour en savoir plus sur cette fonctionnalité, voir les guides d'utilisateur SAP Predictive Analytics pour Data Manager.

(1) Toutes les versions de Teradata

(2) Toutes les versions à partir d'Oracle 10

> **i** Remarque
>
> Pour modifier les paramètres du mode Explain, contactez votre administrateur SGDB qui se reportera au document *Mode Explain* disponible dans la section *Documents de support et d'intégration* de la documentation SAP Predictive Analytics.

1. Dans la liste déroulante *Champ identifiant*, sélectionnez la colonne contenant l'identifiant de l'entité.
2. Facultatif : pour filtrer l'entité de façon à supprimer certaines lignes en fonction d'une condition spécifique, cliquez sur le bouton *Modifier les filtres*. L'existence d'un filtre est indiquée comme ci-dessous :
   - : aucun filtre n'a été défini
   - : un filtre existe déjà
3. Cliquez sur le bouton *Suivant* pour valider la création ou les modifications de l'entité.
4. Si une entité portant le même nom existe déjà, une boite de dialogue vous demande de préciser le comportement à adopter :

- Cliquez sur *Oui* pour enregistrer l'entité avec un nouveau numéro de version.
- Cliquer sur *Non* si vous voulez écraser l'entité existante.

4.15 Modification des filtres

Cet onglet vous permet de créer et d'appliquer des filtres sur les données pour ne conserver que les enregistrements qui vous intéressent.

- Pour définir un filtre
  1. Dans la section *Ne conserver que les entrées qui rendent vrai pour*, choisissez si vous voulez conserver les enregistrements répondant à toutes les conditions ou seulement à au moins l'une d'elles.
  2. Cliquez sur la flèche située à droite du bouton *Ajouter*. L'éditeur de formule s'ouvre, vous permettant ainsi de spécifier une nouvelle condition.
- Pour modifier une condition
  1. Dans la *Liste des conditions du filtre*, sélectionnez la condition que vous voulez modifier.
  2. Cliquez sur le bouton *Modifier*.
- Pour supprimer une condition de filtre
  1. Dans la *Liste des conditions du filtre*, sélectionnez la condition que vous voulez supprimer.
  2. Cliquez sur le bouton *Supprimer*.

Comment utiliser l'éditeur de formule

L'éditeur de formule vous permet de créer des champs (un par un ou en masse) et de modifier des conditions comme vous le feriez avec une calculatrice.

Création d'un champ

- Pour créer un champ :
  1. Entrez une formule dans la zone de texte qui se trouve dans la partie supérieure du panneau. Vous pouvez utiliser les fonctions d'aide à la saisie suivantes :
     - Cliquez deux fois sur le nom d'une fonction située dans le cadre *Fonctions* ou d'une variable (champ ou invite ou ensemble de champs prédéfini) située dans le cadre *Variables* pour l'insérer à l'endroit du curseur. Dans le cas d'une fonction, un patron de la fonction avec des informations sur les paramètres à renseigner est inséré.

- Vous pouvez également insérer ces éléments à la position voulue en utilisant le glisser-déposer depuis l'un des arbres vers la zone de texte.
- Lors de la saisie d'un nom de variable (champ, invite ou ensemble), appuyez simultanément sur les touches `Ctrl` + `Espace` pour afficher une liste des noms de variable commençant par le texte saisi.
  La couleur de la pastille située au-dessus de la zone *Messages* vous informe de l'état de la formule.

	Si la pastille est...	la formule...
	rouge	contient une erreur, qui est signalée dans la zone des messages. Il n'est pas possible de la valider (le bouton Suivant est désactivé).
	jaune	peut être validée mais certaines incohérences sont possibles et sont signalées dans la zone des messages.
	verte	est valide.

  Dans le cas d'une erreur ou d'un avertissement, la zone *Message* située dans la partie inférieure du panneau vous donne des détails pour vous aider à comprendre la raison de l'erreur ou de l'avertissement.

2. Pour valider définitivement la formule, cliquez sur le bouton *Suivant*.
3. Saisissez un nom pour le nouveau champ sous *Nom*.
4. Cliquez sur le bouton *OK*.

Fonctionnalités avancées

Ensembles de champs

Il est courant de vouloir créer plusieurs champs en appliquant un même calcul à différent champs existants. L'éditeur de formule vous permet de faire cela grâce à l'utilisation des ensembles de champs.

Par exemple, l'utilisation des ensembles de champs vous permet d'additionner un grand nombre de champs, ou de calculer leur maximum.

Ensembles de champs nommés

- Pour créer un ensemble de champs nommé :
  1. Dans la section *Variables*, cliquez deux fois sur l'option *Ensemble de champs*. Un sous-arbre s'affiche, listant les ensembles de champs existants.
  2. Cliquez deux fois sur l'option *Créer un ensemble de champs...* Une fenêtre s'affiche listant tous les champs disponibles dans votre table d'analyse.
  3. Dans le champ *Masque d'alias*, saisissez un masque permettant de filtrer les champs par leur nom. Un masque est constitué d'une partie commune au nom de tous les champs que vous voulez voir s'afficher dans la liste et du caractère étoile (*) qui permet de compléter les parties de nom qui diffèrent selon les champs. L'étoile peut être utilisée en début, milieu ou fin de mot et autant de fois que nécessaire.
  4. Décochez les champs que vous ne souhaitez pas conserver dans l'ensemble de champs.
  5. Dans le champ *Nom*, saisissez le nom du nouvel ensemble de champs.

6. Cliquez sur le bouton OK. La fenêtre se ferme et le nouvel ensemble de champs apparaît dans la liste sous l'élément *Ensemble de champs*.

- Pour modifier un ensemble de champs nommé :
  1. Cliquez sur l'ensemble de champs que vous souhaitez modifier pour le sélectionner.
  2. Cliquez avec le bouton droit de la souris sur l'ensemble de champs sélectionné. Un menu contextuel s'affiche.
  3. Cliquez sur l'option Modifier. La fenêtre d'édition des ensembles de champs s'affiche.

     > **i** Remarque
     >
     > Modifier le nom d'un ensemble de champs revient à le dupliquer.

- Pour supprimer un ensemble de champs nommé :
  1. Cliquez sur l'ensemble de champs que vous souhaitez modifier pour le sélectionner.
  2. Cliquez avec le bouton droit de la souris sur l'ensemble de champs sélectionné. Un menu contextuel s'affiche.
  3. Cliquez sur l'option *Supprimer*.

Ensembles de champ créés à la volée

Lorsque vous ne souhaitez pas appliquer un calcul sur les champs dont les noms ont une racine commune, vous pouvez créer un ensemble de champs à la volée dans le champ de texte de la formule.

Un ensemble de champ créé à la volée est défini par un **masque**, correspondant à une partie fixe commune aux noms de tous les champs souhaités et à un caractère générique représentant la partie du nom qui change pour chaque champ. Trois caractères génériques peuvent être utilisés pour définir les ensembles de champs : le signe arobase (@), le signe dièse (#) et le signe du dollar ($). Seul un caractère générique peut être utilisé pour chaque ensemble de champ, et un même caractère générique ne peut pas être utilisé deux fois dans la même formule.

- Pour créer un ensemble de champs à la volée :
  - Dans le champ de texte de la formule, saisissez le masque correspondant aux champs auxquels vous voulez ajouter un calcul, par exemple income_@, puis utilisez-le comme champ standard.
- Pour utiliser plusieurs ensembles de champs :
  1. Saisissez une formule à l'aide des ensembles de champs tel que défini précédemment. La zone *Messages* située dans la partie inférieure du panneau indique le nombre de champs allant être créés.
  2. Lorsque vous utilisez plusieurs ensembles de champs dans la même formule, vous devez sélectionner dans la liste déroulante *Association des champs* la manière dont les champs de ces ensembles seront associés.
     - *Associer par position* : les champs de chaque ensemble sont associés en fonction de leur position dans la table de base de données.
     - *Associer par valeur* : les champs de chaque ensemble sont associés en fonction de la valeur représentée par les caractères génériques utilisés pour définir les ensembles de champ.
     - *Effectuer le produit cartésien* : tous les champs d'un ensemble sont associés à tous les champs d'un autre ensemble.

     Par exemple : prenons le cas d'une table de base de données contenant les champs suivants organisés comme suit :
     income_january
     income_april
     income_february

income_march
expenses_march
expenses_january
expenses_april
La formule suivante `income_@ - expenses_#` utilise deux ensembles de champs, un regroupant tous les champs `income_` et le deuxième avec tous les champs `expenses_` :

income_@	expenses_#
income_january	expenses_march
income_april	expenses_january
income_february	expenses_april
income_march	

L'option *Associer par position* entraîne les calculs suivants :

Calcul	Position
income_january - expenses_march	1
income_april - expenses_january	2
income_february - expenses_april	3

Si l'option *Associer par valeur* est sélectionnée, l'application essayera d'associer la valeur représentée par @ et la valeur représentée par `#` , entraînant les calculs suivants :

Calcul	Valeurs @ et #
income_january - expenses_january	january
income_april - expenses_april	april
income_march - expenses_march	march

L'option *Effectuer le produit cartésien* entraîne les calculs suivants :

Calcul
income_january - expenses_march
income_january - expenses_january
income_january - expenses_april

Calcul
income_april- expenses_march
income_april - expenses_january
income_april - expenses_april
income_february - expenses_march
income_february - expenses_january
income_february - expenses_april
income_march - expenses_march
income_march - expenses_january
income_march - expenses_april

3. Cliquez sur *Suivant* pour valider la création des champs. Une boîte de dialogue vous permettant de nommer les nouveaux champs s'ouvre.
4. Saisissez une racine commune à tous les champs venant d'être créés.
5. Si la formule utilise uniquement les ensembles de champ nommés ou si vous n'avez pas besoin de contrôler davantage l'appellation, passez à l'étape suivante. Si vous avez utilisé un ou plusieurs caractères génériques, vous pouvez les utiliser dans l'alias pour créer un modèle d'appellation. Les caractères génériques sont remplacés par les parties de nom correspondant.
6. Cliquez sur *OK* pour créer les champs.

Utilisation d'un ensemble de champs en tant que liste de variables pour une fonction n-aire.

Un ensemble de champs peut être utilisé pour définir une liste des arguments pour des fonctions n-aire, c'est-à-dire des fonctions qui comportent un nombre indéfini d'arguments. Lorsqu'un ensemble de champs est utilisé en tant qu'argument dans une fonction n-aire, et dans ce cas uniquement, vous devez encadrer le nom de l'ensemble avec des accolades **{ }** pour qu'il soit interprété comme une liste d'arguments.

Prenons par exemple le cas d'une table de base de données sont les champs income_april, income_may, income_june contiennent le chiffre d'affaires mensuel. Si vous souhaitez créer un champ contenant le chiffre d'affaire mensuel le plus élevé du trimestre, vous devez utiliser la formule suivante `greatestN({income_@})` qui utilise l'ensemble de champs `income_@` en tant que liste de variables, ce qui équivaut à la formule `greatestN(income_april ,income_may ,income_june)`. Toutefois, lorsque vous utilisez l'ensemble de champs tel quel dans la formule `greatestN(income_@)`, trois champs sont créés par les trois formules suivantes : `greatestN(income_april)`, `greatestN(income_may)` et `greatestN(income_june)`, qui ne peut être utilisé.

- Pour utiliser un ensemble de champs en tant que liste de variable :
  1. Créez un ensemble de champs (nommé ou à la volée) tel que décrit précédemment.
  2. Dans le champ de texte de la formule, saisissez la formule contenant une fonction n-aire.
  3. Saisissez le nom de l'ensemble de champs entre accolades `{ }` pour l'utiliser en tant qu'argument pour la fonction. Par exemple, la formule `greatestN({incomes})` créera un champ dont la valeur est la valeur la plus élevée des champs contenus dans l'ensemble de champs appelé incomes.

4.16 Modification des filtres

Cet onglet vous permet de créer et d'appliquer des filtres sur les données pour ne conserver que les enregistrements qui vous intéressent.

- Pour définir un filtre
  1. Dans la section *Ne conserver que les entrées qui rendent vrai pour*, choisissez si vous voulez conserver les enregistrements répondant à toutes les conditions ou seulement à au moins l'une d'elles.
  2. Cliquez sur la flèche située à droite du bouton *Ajouter*. L'éditeur de formule s'ouvre, vous permettant ainsi de spécifier une nouvelle condition.
- Pour modifier une condition
  1. Dans la *Liste des conditions du filtre*, sélectionnez la condition que vous voulez modifier.
  2. Cliquez sur le bouton *Modifier*.
- Pour supprimer une condition de filtre
  1. Dans la *Liste des conditions du filtre*, sélectionnez la condition que vous voulez supprimer.
  2. Cliquez sur le bouton *Supprimer*.

4.17 Agrégats glissants

Ce panneau vous permet de définir des agrégats glissants, c'est-à-dire des agrégats qui seront calculés pour chaque date de l'indicateur de performance sur une période de temps spécifiée.

4.18 Composition de populations estampillées

Ce panneau vous permet de créer une population estampillée en utilisant les opérateurs ensemblistes (union, intersection et différence) sur des populations estampillées existantes.

Définition

Une population estampillée est une liste de paires `<identifiers; time stamps>` : qui représentent des instantanés des entités à un moment donné. Une entité peut figurer plusieurs fois dans une même population estampillée mais associée à des moments différents.

Création ou modification d'une composition

La composition est représentée par un arbre dont le premier niveau est un opérateur ensembliste et les niveaux suivants des populations estampillées ou des opérateurs.

L'icône...	Ayant l'étiquette...	Représente...
	Indéfini	un paramètre non défini
	Concaténation	l'opérateur d'union
	Intersection	l'opérateur d'intersection
	Différence	l'opérateur de différence
	<Nom d'une population estampillée>	la population estampillée nommée

Par défaut, un modèle dont les paramètres ne sont pas définis est affiché.

- Pour créer une composition
  1. Dans le champ *Nom*, saisissez le nom de la nouvelle population estampillée.
  2. Dans le champ *Description*, saisissez une description de la population estampillée.
  3. Cliquez sur le paramètre non défini en tête de liste et sélectionnez le premier opérateur dans le menu affiché.

Nom de l'opérateur	Résultat
Concaténation	l'union des populations estampillées sélectionnées
Intersection	l'intersection des populations estampillées sélectionnées
Différence	la différence entre les populations estampillées sélectionnées

     Le drapeau est remplacé par l'opérateur sélectionné suivi de son nom.

  4. Utilisez les paramètres non définis situés au niveau inférieur pour sélectionner soit une population estampillé sur laquelle appliquer l'opérateur, soit un nouvel opérateur. Si vous sélectionnez un opérateur, un niveau supplémentaire s'affiche vous permettant de paramétrer ses opérandes.
  5. Une fois que tous les paramètres de la composition sont saisis, cliquez sur le bouton *Suivant* pour l'enregistrer.

4.19 Produit cartésien : population estampillée

Le panneau *Produit cartésien* vous permet de créer une population estampillée en croisant une entité et une table de dates, c'est-à-dire en associant chaque ligne de l'entité à chaque date de la table de dates. Dans l'exemple suivant, si la table *T_entité* est croisée avec la table *T_date*, le résultat sera la table *T_population*.

T_référence	T_date	T_résultat	
ref1	date1	ref1	date1
ref2	date2	ref1	date2
ref3	date3	ref1	date3
	date4	ref1	date4
		ref2	date1
		ref2	date2
		ref2	date3
		ref2	date4
		ref3	date1
		ref3	date2
		ref3	date3
		ref3	date4

> **i** Remarque
>
> En fonction de la taille des tables qui seront croisées, cette opération peut générer un grand nombre de lignes et prendre beaucoup de temps.

Définition

Une population estampillée est une liste de paires `<identifiers; time stamps>` : qui représentent des instantanés des entités à un moment donné. Une entité peut figurer plusieurs fois dans une même population estampillée mais associée à des moments différents.

Création d'une population estampillée à partir d'un produit cartésien

- Pour créer une population estampillée grâce à un produit cartésien
  1. Dans le champ *Nom*, saisissez le nom de la population estampillée à créer.
  2. Saisissez une description dans le champ *Description*.
  3. Dans la liste *Table de référence*, sélectionnez l'entité à utiliser.
  4. Dans la liste *Table de date*, sélectionnez la table de date à utiliser. La table de date *ne doit contenir que les dates nécessaires* à la création de la population estampillée. Si cette table de date filtrée n'existe pas, vous pouvez la créer en cliquant sur le bouton *Créer...* situé à droite de la liste *Table de date*. L'éditeur de jeu de données d'analyse s'ouvre alors et vous permet de créer une table correspondant à vos besoins. La table nouvellement créée est sélectionnée par défaut dans la liste *Table de date*.
  5. Dans la liste *Colonne de date*, sélectionnez la date à utiliser dans la population estampillée.
  6. Vous pouvez voir un aperçu de la population estampillée créée en cliquant sur le bouton *Aperçu...*
  7. Cliquez sur le bouton *Suivant* pour créer la population estampillée.

4.20 Population estampillée filtrée

Utilisez ce panneau pour créer ou modifier une population estampillée.

Définition

Une population estampillée est une liste de paires `<identifiers; time stamps>` : la signification sémantique d'une telle construction peut être associée aux instantanés des entités et à un moment donné. De manière générale, une entité donnée peut figurer dans une même population estampillée à des moments différents.

Création ou modification d'une population estampillée

- Pour créer ou modifier une population estampillée
  1. Dans le champ *Nom*, saisissez le nom de la nouvelle population estampillée.
  2. Dans le champ *Description*, saisissez une description.
- Paramètres
  1. Dans la liste déroulante *Basée sur l'entité*, sélectionnez l'entité sur laquelle sera construite la population estampillée.
  2. Facultatif : pour filtrer la population estampillée de façon à supprimer certaines lignes en fonction d'une condition spécifique, cliquez sur le bouton *Modifier les filtres*. L'existence d'un filtre est indiquée comme ci-dessous :
     - : aucun filtre n'a été défini
     - : un filtre existe déjà
  3. Dans la section *Horodatage*, sélectionnez la date à utiliser pour la population estampillée. Il peut s'agir d'un champ (calculé ou non), d'une constante ou d'une invite.
  4. Cliquez sur *Modifier le stockage de KxTimeStamp* pour sélectionner le stockage de votre choix.
     Sélectionnez *Date* si vous souhaitez que le format de stockage soit *AAAA-MM-JJ*.
     Sélectionnez *Date avec horaire* si vous souhaitez que le format de stockage soit *AAAA-MM-JJ HH:MM:SS*.
  5. Cliquez *OK* pour valider.
- Cible
  1. Dans la liste *Cible*, sélectionnez la cible que vous souhaitez utiliser. Si vous ne souhaitez pas utiliser une cible, sélectionnez *Ne pas définir de cible pour l'instant* et accédez directement à la section *Pour valider la création ou la modification d'une population estampillée*.
  2. Si vous avez sélectionné une cible et que vous souhaitez accélérer le processus d'entraînement, vous pouvez cocher la case *Activer l'échantillonnage stratifié*. Lors de l'activation de l'échantillonnage stratifié, le nombre de lignes dans le jeu de données est réduit mais la quantité d'information est conservée. Cela peut s'avérer utile lorsqu'il y a peu de cas positifs à modéliser. Par exemple, si un jeu de données contient 1 000 cas positifs et 1 000 000 cas négatifs, l'activation de l'échantillonnage stratifié entraîne la conservation des 1 000 cas positifs et l'échantillonnage des cas négatifs. Dans

cet exemple, si nous conservons 1 cas négatif sur 10, le jeu de données contiendra 1 000 cas positifs et 100 000 cas négatifs : la proportion cible est de 1 % alors qu'elle était de 0,01 % dans le jeu de données d'origine. Comme le jeu de données est environ dix fois plus petit, l'entraînement sera beaucoup plus rapide.

3. Cochez la case *Activer les poids* pour générer une colonne de poids permettant d'associer chaque cas négatif avec un poids. Dans l'exemple ci-dessus, vous définissez le poids sur 10. Comme un poids de dix représente approximativement dix cas négatifs, cela permet de conserver des statistiques qui représentent la population initiale sans échantillonnage et donc d'avoir des mesures de la moyenne, des effectifs pondérés et de la capacité prédictive (KI).
4. Cliquez sur le bouton *Définir les proportions initiales* pour extraire ou définir manuellement les catégories.

- Pour valider la création ou la modification d'une population estampillée
  1. Cliquez sur le bouton *Suivant* pour valider la création ou la modification de la population estampillée.
  2. Si une population estampillée portant le même nom existe déjà, une boîte de dialogue vous demande de préciser le comportement à adopter :
     - Cliquez sur *Oui* pour enregistrer la population estampillée avec un nouveau numéro de version.
     - Cliquer sur *Non* si vous voulez écraser la population estampillée existante.

4.21 Produit cartésien

Le panneau *Produit cartésien* vous permet de créer une manipulation de données en croisant une table de références et une table de dates, c'est-à-dire en associant chaque ligne de la table de références à chaque date de la table de date. Dans l'exemple suivant, si la table *T_référence* est croisée avec la table *T_date*, le résultat sera la table *T_résultat*.

T_référence	X	T_date	=>	T_résultat
ref1	date1	ref1	date1	
ref2	date2	ref1	date2	
ref3	date3	ref1	date3	
	date4	ref1	date4	
		ref2	date1	
		ref2	date2	
		ref2	date3	
		ref2	date4	
		ref3	date1	
		ref3	date2	
		ref3	date3	
		ref3	date4	

> **i** Remarque
>
> En fonction de la taille des tables qui seront croisées, cette opération peut générer un grand nombre de lignes et prendre beaucoup de temps.

- Pour créer une manipulation de données grâce à un produit cartésien :
  1. Dans le champ *Nom*, saisissez le nom de la manipulation de données à créer.
  2. Saisissez une description dans le champ *Description*.
  3. Dans la liste *Table de référence*, sélectionnez la table de références à utiliser.
  4. Dans la liste *Table de date*, sélectionnez la table de date à utiliser. La table de date *ne doit contenir que les dates nécessaires* à la création de la manipulation de données. Si cette table de date filtrée n'existe pas, vous pouvez la créer en cliquant sur le bouton *Créer...* situé à droite de la liste *Table de date*. L'éditeur de jeu de données d'analyse s'ouvre alors et vous permet de créer une table correspondant à vos besoins. La table nouvellement créée est sélectionnée par défaut dans la liste *Table de date*.
  5. Dans la liste *Colonne de date*, sélectionnez la date à utiliser dans la manipulation de données.
  6. Vous pouvez voir un aperçu de la manipulation de données créée en cliquant sur le bouton *Aperçu...*
  7. Cliquez sur le bouton *Suivant* pour créer la manipulation de données.

4.22 Ajout d'un filtre dans un jeu de données

Vous avez la possibilité d'appliquer un filtre à votre jeu de données afin d'accélérer le processus d'apprentissage et d'optimiser le modèle qui en résulte.

Prenons l'exemple dans ce qui suit de l'élaboration du filtre complexe suivant :

```
( (age>32) OR (native-country<>"United-States") ) AND ( (marital-
status<>"Divorced") AND (hours-per-week<45) )
```


- Ajout d'une condition [page 66]
- Ajout d'une conjonction logique [page 67]
- Changement de l'ordre [page 68]
- Suppression d'un nœud [page 68]
- Aperçu des données [page 68]
- Enregistrement du filtre [page 68]
- Chargement d'un filtre [page 69]

4.22.1 Ajout d'une condition

Le filtre est représenté par une arborescence. Les conditions sont reliées entre elles par la conjonction logique située au-dessus dans l'arborescence. Pour créer le filtre complexe de l'exemple précédent, ajoutez les conditions `(age>32)` et `(native-country!=united-states))` à la conjonction logique `Or` créée à l'étape

précédente et les conditions `(marital-status!=divorced)` et `(hours-per-week<45)` à la conjonction logique And créée à l'étape précédente.

Le filtre complexe conservera tous les enregistrements dans le jeu de données pour lesquels l'âge est supérieur à 32 **ou** le pays de naissance n'est pas les États-Unis d'une part, et pour lesquels la situation de famille n'est pas "divorcé" And l'emploi du temps hebdomadaire est inférieur à 45 d'autre part. Les enregistrements conservés sont *à l'intersection* (And) de ces deux jeux d'enregistrements.

- Pour ajouter une condition
  1. Cliquez sur le nœud auquel vous souhaitez ajouter une condition. La nouvelle condition sera ajoutée comme une sous-arborescence du nœud actuellement sélectionné.
  2. Cliquez sur le bouton *Ajouter une condition*.
     La fenêtre *Définir une condition* s'ouvre.
  3. Choisissez une variable dans la première liste déroulante.
  4. Choisissez un opérateur dans la deuxième liste.
  5. Indiquez une valeur dans la troisième liste :
     - Pour une variable du type "number" entrez une valeur.
     - Pour une variable du type "string" choisissez une valeur dans la liste. Si cette liste est vide, cliquez sur le bouton pour extraire les catégories.
  6. Cliquez sur *OK*.

> **i** Remarque
>
> Vous pouvez modifier une condition en cliquant deux fois dessus.

4.22.2 Ajout d'une conjonction logique

Les conjonctions logiques ou nœuds s'appliquent aux éléments situés en dessous. Pour créer le filtre complexe ci-dessus, il faut conserver le nœud And existant, il correspond à l'opérateur entre les deux expressions principales entre parenthèses. Vous devez ensuite créer deux nœuds correspondant aux opérateurs dans chaque expression entre parenthèses : un nœud `Or` pour la première expression et un nœud And pour la seconde.

- Pour ajouter une conjonction logique :
  1. Cliquez sur le nœud auquel vous souhaitez ajouter une conjonction logique. La nouvelle conjonction sera ajoutée comme une sous-arborescence du nœud actuellement sélectionné.
  2. Cliquez sur le bouton *Ajouter un "ET" logique* ou sur le bouton *Ajouter un "OU" logique*.

> **i** Remarque
>
> Vous pouvez modifiez le type de conjonction en cliquant deux fois dessus.

4.22.3 Changement de l'ordre

Vous pouvez changer l'ordre :

- Des nœuds (ou conjonctions logiques) pour accélérer l'application du filtre en mettant les nœuds dont les conditions qui ont une grande probabilité de s'avérer fausses en haut de la liste.
- Des conditions au sein d'un nœud en mettant les conditions qui ont une grande probabilité de s'avérer fausses en haut de la liste.
- Pour changer l'ordre :
  1. Sélectionnez le nœud ou la condition que vous voulez déplacer vers le haut ou vers le bas.
  2. Utilisez les boutons ^ et v pour changer sa position dans le filtre.

4.22.4 Suppression d'un nœud

- Pour supprimer un nœud :
  1. Sélectionnez le nœud que vous voulez supprimer.
  2. Cliquez sur le bouton *Supprimer le nœud sélectionné*.

4.22.5 Aperçu des données

Vous pouvez visualiser le jeu de données que vous obtiendrez en appliquant le filtre.

- Pour afficher le jeu de données filtré
  - Cliquez sur le bouton *Voir les données*.
    Une fenêtre contextuelle s'ouvre, affichant tous les enregistrements du jeu de données qui satisfont les conditions du filtre configuré.

4.22.6 Enregistrement du filtre

Vous pouvez enregistrer le filtre créé pour le réutiliser ultérieurement sans être obligé de recréer un filtre avec les mêmes conditions.

- Pour enregistrer un filtre :
  1. Cliquez sur *Enregistrer ce filtre*.
     La fenêtre *Enregistrer ce filtre* s'affiche.
  2. Dans la liste *Type de données*, sélectionnez le format de l'enregistrement.
  3. Utilisez le bouton *Parcourir* à droite du champ *Dossier* pour choisir un dossier ou une base de données pour l'enregistrement.
  4. Dans le champ *Fichier*, entrez le nom du fichier ou de la table pour l'enregistrement du filtre.

5. Cliquez sur le bouton *OK*.

4.22.7 Chargement d'un filtre

Pour filtrer un jeu de données, vous pouvez utiliser un filtre préalablement créé et enregistré pour ce jeu de données.

- Pour charger un filtre existant :
  1. Cliquez sur le bouton *Charger un filtre existant*.
  2. Utilisez la liste déroulante *Type de données* pour sélectionner le format du filtre.
  3. Utilisez le bouton *Parcourir* à droite du champ *Dossier* pour choisir le dossier ou la base de données où doit être stocké le filtre.
  4. Utilisez le bouton *Parcourir* à droite du champ *Fichier* pour choisir le fichier ou la table contenant le filtre.
  5. Cliquez sur le bouton *OK*.

4.22.8 Échantillonnage

Avec le filtre *Échantillonnage*, vous réduisez le nombre de lignes de votre jeu de données. Vous pouvez conserver une plage de lignes ou demander à SAP Predictive Analytics de sélectionner des lignes aléatoires.

L'échantillonnage peut être réalisé de la façon suivante :

- soit la sélection des n premières lignes dans le jeu de données d'origine ;
- soit la sélection aléatoire d'une proportion du jeu de données d'origine. Une graine aléatoire permet de sélectionner les mêmes enregistrements dans le jeu de données d'origine lors de l'échantillonnage.

- Pour paramétrer un échantillonnage aléatoire
  1. Cochez l'option *Échantillonnage aléatoire*.
  2. Dans le champ *Graine aléatoire*, saisissez une valeur pour l'échantillonnage. La valeur par défaut de la graine est 1960.
  3. Dans la partie *Proportion de lignes à conserver*, utilisez le curseur pour sélectionner la proportion de l'observation qui doit apparaître dans le jeu de données. La sélection aléatoire génère un index aléatoire pour chaque ligne du jeu de données basé sur la graine. Une même graine pour un même index.
     Vous pouvez affiner la sélection en cochant l'option *Avancé*. Le curseur est remplacé par deux champs qui vous permettent d'indiquer un intervalle correspondant aux index à conserver. Par exemple, si vous sélectionnez [0, 20 %], vous sélectionnez les lignes avec un index compris entre 0 et 0,2. Ainsi [0, 20 %] ne présente pas les mêmes lignes que [80 %, 100 %](index compris entre 0,8 et 1). En réalité, si vous souhaitez apprendre sur 80 % de votre jeu de données, appliquer sur les 20 % restants et effectuer un échantillonnage aléatoire, c'est la seule méthode. Vous pouvez alors créer deux jeux de données complémentaires en conservant les mêmes paramètres de curseur et de graine et en sélectionnant deux intervalles complémentaires.
- Pour paramétrer une sélection par lignes
  1. Cochez l'option *Conserver les lignes de 1 à*.

2. Puis saisissez le numéro de ligne pour sélectionner la plage que vous souhaitez conserver.
3. Une fois l'échantillonnage spécifié, fermez la fenêtre *Paramètres*.

4.23 Source des données d'événement

Les données du journal des événements doivent être sélectionnées dans cet écran.

Un jeu de données événementielles ou transactionnelles doit contenir au minimum :

- une date d'événement comme une date de naissance ou un début d'essai au format AAAA/MM/JJ ;
- un identifiant de référence (par exemple l'identifiant d'un client) utilisé pour joindre les données événementielles ou transactionnelles à la table de clients statique ou de référence définie au préalable.

Ces données peuvent aussi contenir des valeurs numériques (par exemple un montant d'achat) qui seront utilisées pour calculer des agrégats spécifiques.

- Pour sélectionner la source des données événementielles
  1. Utilisez le bouton *Parcourir* pour sélectionner le répertoire du fichier texte ou la source ODBC dans la ligne *Répertoire*.
  2. Sélectionnez le journal des événements ou les données transactionnelles dans la ligne *Événements* en utilisant le bouton *Parcourir*.

4.23.1 Jeu de données d'événement pour la vérification des écarts

Le panneau *Jeu de données d'événement pour la vérification des écarts* vous permet de sélectionner le jeu de données pour lequel vous souhaitez vérifier s'il y a des écarts par rapport au jeu de données utilisé pour la création du modèle.

- Pour sélectionner un jeu de données
  1. Sélectionnez le type de fichier de jeu de données dans la liste *Type de données*.
  2. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de données contenant le jeu de données.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Données* pour sélectionner le fichier ou la table contenant le jeu de données.
  4. Cliquez sur le bouton *Suivant*.

Il est également possible de ne sélectionner qu'une partie du jeu de données grâce à l'échantillonnage.

- Pour échantillonner un jeu de données
  1. Cliquez sur le bouton *Paramètres d'échantillonnage* situé à droite du champ *Jeu de données*.
     Une fenêtre s'ouvre vous permettant de saisir les paramètres d'échantillonnage.
  2. Cochez l'option *Voulez-vous utiliser les options d'échantillonnage ?*

L'échantillonnage peut être réalisé de la façon suivante :

- Sélection aléatoire d'une proportion des données originales. Une graine aléatoire permet de sélectionner les mêmes enregistrements dans les données originales lors de l'échantillonnage.
- Sélection d'une première ligne et d'une dernière ligne spécifiques dans les données originales.
- Combinaison de la sélection par lignes et de la sélection aléatoire.

- Pour paramétrer une sélection par lignes
  1. Cochez l'option *Sélection par lignes* dans la section *Type d'échantillonnage*.
  2. Dans le champ *Première ligne*, saisissez le numéro de la ligne par laquelle commencera le jeu de données échantillonné.
  3. Dans le champ *Dernière ligne*, saisissez le numéro de la ligne par laquelle se terminera le jeu de données échantillonné.
- Pour paramétrer un échantillonnage aléatoire
  1. Cochez l'option *Sélection aléatoire* dans la section *Type d'échantillonnage*.
  2. Dans la section *Sélection aléatoire*, utilisez le curseur *Proportion* pour sélectionner le pourcentage du jeu de données d'origine que vous souhaitez conserver.
  3. Dans le champ *Graine aléatoire*, saisissez une valeur pour l'échantillonnage. Cette étape est facultative.
  4. Utilisez le curseur *Proportion* pour sélectionner le pourcentage du jeu de données d'origine que vous souhaitez conserver. Une fois le curseur positionné, vous pouvez sélectionner la partie haute ou la partie basse en cochant la case d'option correspondant. Cette option associée à l'utilisation de la même valeur de départ permet de créer deux jeux de données complètement différents à partir de celui d'origine.

4.23.2 Jeu de données d'événements pour la simulation

Ce panneau permet de spécifier un jeu de données d'événements à utiliser pour simuler un modèle sur la base des fonctionnalités de l'analyseur de séquences et de la journalisation des événements.

Utilisez le bouton *Parcourir* pour sélectionner le répertoire du fichier texte ou la source ODBC dans la ligne *Répertoire*. Sélectionnez ensuite le journal des événements ou les données transactionnelles dans la ligne *Événements* à l'aide du bouton *Parcourir*.

Un jeu de données événementielles ou transactionnelles doit contenir au minimum :

- une date d'événement comme une date de naissance ou un début d'essai au format AAAA/MM/JJ ;
- un identifiant de référence (par exemple l'identifiant d'un client) utilisé pour joindre les données événementielles ou transactionnelles à la table de clients statique ou de référence définie au préalable.

Ces données peuvent aussi contenir des valeurs numériques (par exemple un montant d'achat) qui seront utilisées pour calculer des agrégats spécifiques.

Lorsque vous simulez un modèle utilisant la journalisation des événements, vous devez spécifier la *date de référence* à utiliser dans la section *Fixe* située en bas du panneau.

La *date de référence* est la date à partir de laquelle vont être définies les périodes à analyser. Cette date peut être :

- soit *fixe*, c'est à dire une date spécifique, par exemple 24-05-2004,
- soit *variable*, c'est à dire une date définie par une des variables du jeu de données. Par exemple, la date du premier achat, la date de résiliation d'un abonnement, etc.

1. Sélectionnez le type de la *Date de référence* (Fixe ou Variable).
2. Selon l'option choisie, spécifiez une date fixe dans le champ *YYYY-MM-DD hh:mm:ss*, ou sélectionnez une des variables date de la table de référence dans le menu déroulant *Colonne*.

4.24 Description des données d'événement

Pour que vos données soient interprétables et analysables par les fonctions SAP Predictive Analytics, elles doivent être décrites. En d'autres termes, le fichier de description doit spécifier la nature de chaque variable en déterminant leur format de stockage et leur type.

- Définitions [page 72]
- Description des données [page 74]
- Enregistrement d'une description [page 75]

4.24.1 Définitions

Description d'une variable

La description précise d'une variable permet aux fonctionnalités d'Automated Analytics d'effectuer les calculs appropriés lors de la génération d'un modèle en vue d'obtenir les meilleurs résultats.

Chaque variable ou attribut est décrit par les informations suivantes :

Champ	Description	Valeurs
Nom	nom de la variable (ne peut être modifié)	

Champ	Description	Valeurs
Type	type de valeurs enregistrées dans cette variable Remarque : Lorsque vous déclarez une variable comme date ou date et heure, codeur de date extrait automatiquement de cette variable les informations de date telles que le jour du mois, l'année, le trimestre, etc. Des variables supplémentaires contenant ces informations sont créées lors de la génération et utilisées comme variables d'entrée du modèle. Le codeur de date est désactivé pour les modèles de série temporelle.	*Nombre* : nombres "calculables" (attention : les numéros de téléphone ou de compte ne doivent pas être considérés comme des nombres). *Entier* : nombres entiers *Chaîne* : chaînes de caractères *Date et heure* : date et heure *Date* : dates
Valeur	décrit le type de valeur Attention : lorsque vous créez un modèle d'encodage textuel, il faut que le jeu de données contienne au moins une variable textuelle pour pouvoir passer au panneau suivant.	*Continue* : variable numérique à partir de laquelle une moyenne, une variance, etc. peuvent être calculées *Nominale* : variable de catégorie (seule valeur possible pour une chaîne) *Ordinale* : variable numérique discrète dont l'ordre relatif est important Textuelle : variable textuelle contenant des expressions, des phrases ou des textes complets
Clé	variable clé ou identifiant pour l'enregistrement Il est possible de déclarer des clés multiples qui seront construites selon l'ordre indiqué (1-2-3-...).	*0* : la variable n'est pas un identifiant, *1* : identifiant primaire, *2* : identifiant secondaire, ...
Ordre	indique si cette variable représente un ordre naturel. Si la valeur est 1, la variable est utilisée dans les expressions SQL dans une condition "ORDER BY". Avertissement : définir une variable comme variable d'ordre aura une incidence sur le choix d'une stratégie de partition. Seules les stratégies séquentielles peuvent être utilisées avec une variable d'ordre.	*0* : la variable ne représente pas un ordre naturel *1* : la variable représente un ordre naturel
Inconnu	Chaîne utilisée dans les données pour représenter une valeur manquante	Par exemple : 999 ou #Vide

Champ	Description	Valeurs
Groupe	indique le nom du groupe auquel appartient cette variable. Les variables d'un même groupe contiennent une même information et donc ne seront pas croisées pour les modèles d'un degré supérieur à 1. Ce paramètre sera utilisable dans une version future.	
Description	étiquette de description supplémentaire pour la variable	

Bibliothèque de variables

La bibliothèque de variables vous permet de stocker les descriptions de variables que vous avez déjà utilisées. Vous pourrez ainsi les réutiliser automatiquement lors de la description d'un jeu de données par analyse.

4.24.2 Description des données

Analyse du jeu de données

L'analyse du jeu de données d'entraînement vous permettra de commencer avec une description obtenue soit par l'analyse des cents premières lignes de données, soit par l'utilisation d'une description issue de la bibliothèque de variables si une variable de même nom s'y trouve. Vous pouvez ensuite modifier les descriptions proposées (en vérifiant surtout le genre des variables).

> **i** Remarque
>
> Pour les fichiers texte, il est bon d'avoir une première ligne d'en-tête contenant les noms des variables.

Lorsque vous créez la description par *analyse* automatique, vérifiez chaque variable afin de valider la description comme indiqué dans la définition Description d'une variable.

- Pour analyser le jeu de données
  - Cliquez sur le bouton *Analyser* situé dans l'onglet *Accueil* du ruban.
    Les variables dont la description provient de la bibliothèque de variables sont signalées par l'icône dans la colonne *Index*.

Chargement d'une description du jeu de données

Pour décrire le jeu de données, vous pouvez utiliser un fichier ou une table de base de données défini(e) précédemment et contenant les descriptions de toutes les variables.

- Pour ouvrir une description de jeu de données
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Ouvrir une description*.
  2. Dans la fenêtre *Ouvrir une description pour...*, sélectionnez le format de la description dans la liste déroulante *Type de données*.
  3. Utilisez le bouton *Parcourir* à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où doit être stockée la description.
  4. Utilisez le bouton *Parcourir* situé à droite du champ *Description* pour sélectionner le fichier ou la table contenant la description.
  5. Cliquez sur *OK*. La fenêtre se ferme et la description s'affiche dans le panneau principal. Vous pouvez alors la modifier si nécessaire.

4.24.3 Enregistrement d'une description

Dans la bibliothèque de variables

Vous pouvez choisir d'enregistrer les descriptions de toutes les variables, ou sélectionner uniquement celles qui vous intéressent. Vous pouvez aussi supprimer des variables de la bibliothèque de variables.

Si vous avez défini un référentiel de métadonnées unique (option accessible via le bouton *Aide*, puis *Modifier les options...*, onglet *Référentiel de métadonnées*), les descriptions sont enregistrées dans ce référentiel sinon elles sont enregistrées au même emplacement que le jeu de données que vous décrivez.

- Pour enregistrer toutes les descriptions :
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Enregistrer dans la bibliothèque de variables*.
  2. Sélectionnez l'option *Toutes les variables*.
- Pour enregistrer uniquement certaines descriptions :
  1. Sélectionnez la variable dont vous voulez enregistrer la description. Pour sélectionner plusieurs variables, utilisez la touche `Ctrl`.
  2. Cliquez sur le bouton *Enregistrer dans la bibliothèque de variables*.
  3. Sélectionnez l'option *Variables sélectionnées*.
- Pour supprimer de la bibliothèque de variables :
  1. Sélectionnez la variable à supprimer de la bibliothèque de variables et cliquez sur *Supprimer* de la bibliothèque de variables.

Dans un fichier ou une table de base de données

Vous pouvez enregistrer l'ensemble des descriptions du jeu de données dans un fichier ou dans une table de base de données.

- Pour enregistrer la description du jeu de données dans un fichier ou une table :
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Enregistrer la description*. La fenêtre *Enregistrer une description de...* s'affiche.
  2. Dans la liste *Type de données*, sélectionnez le format dans lequel vous voulez enregistrer la description.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour choisir le dossier ou la base de données où vous souhaitez enregistrer la description.
  4. Dans le champ *Description*, saisissez le nom du fichier ou de la table où vous souhaitez enregistrer la description.
  5. Cliquez sur *OK*.

4.25 Application de données événementielles

Cet écran permet à l'utilisateur de sélectionner un nouveau jeu de données d'événement lors de l'analyse des séquences d'événements, de la journalisation d'événements ou de l'application d'un modèle de règles d'association.

Un jeu de données événementielles ou transactionnelles doit contenir au minimum :

- une date d'événement comme une date de naissance ou un début d'essai au format AAAA/MM/JJ ;
- un identifiant de référence (par exemple l'identifiant d'un client) utilisé pour joindre les données événementielles ou transactionnelles à la table de clients statique ou de référence définie au préalable.

Ces données peuvent aussi contenir des valeurs numériques (par exemple le montant d'achat) qui seront utilisées pour calculer des agrégats spécifiques.

- Pour sélectionner un nouveau jeu de données d'événement
  1. Utilisez le bouton *Parcourir* pour sélectionner le répertoire du fichier texte ou la source ODBC dans la ligne *Répertoire*.
  2. Sélectionnez le journal des événements ou les données transactionnelles dans la ligne *Événements* en utilisant le bouton *Parcourir*.

4.26 Gestion des performances lors de l'utilisation des bases de données

Avant de demander des données stockées dans une base de données (1) Oracle(2) ou SQLServer 2005, l'application utilise une fonctionnalité appelée **mode Explain** qui catégorise les performances des requêtes SQL dans plusieurs classes définies par l'utilisateur. Pour accélérer et alléger la tâche, ce classement est effectué **sans exécuter réellement la requête SQL complète**.

> **i** Remarque
>
> - (1) Pour toutes les versions de Teradata.
> - (2) Pour toutes les versions à partir d'Oracle 10.

L'objectif est de vous permettre d'estimer la charge de la requête SQL avant de l'exécuter, puis de décider, éventuellement en fonction de la politique informatique de l'entreprise, si la requête SQL peut être réellement utilisée.

Une politique informatique d'entreprise peut favoriser l'interactivité et définir 3 classes de requêtes SQL, chacune avec une durée maximum :

- *Immédiate* : *durée < 1 s*. La requête est acceptée et exécutée immédiatement.
- *En masse* : *1 s <= durée < 2 s*. La requête est acceptée et sera exécutée à la prochaine période d'inactivité.
- *Rejetée* : *2 s <= durée*. La requête n'est jamais exécutée.

Le nombre, les noms et les limites de classes sont définis par l'utilisateur afin que ces valeurs correspondent à la configuration SGBD actuelle et aux règles d'utilisation SGBD.

Le mode Explain a été configuré

Si le mode Explain a été configuré par votre administrateur DBMS, deux résultats sont possibles :

- **la requête est acceptée et exécutée** : tout est entièrement transparent. L'application accède aux données sans autre intervention de l'utilisateur.
- **la requête doit être validée avant d'être exécutée** : une fenêtre contextuelle s'ouvre et affiche un message configuré par l'administrateur DBMS. Une requête devant être validée peut être classée de deux manières : **taille moyenne** ou **très grande taille** .

Si la taille de la requête est considérée comme moyenne, vous devrez probablement demander la marche à suivre à votre administrateur :

- Si celui-ci autorise la requête, cliquez sur *Continuer*. La fenêtre contextuelle se ferme et l'action requise se poursuit.
- S'il n'autorise pas la requête, cliquez sur *Arrêter la requête*, la fenêtre contextuelle se ferme, mais aucune action n'est exécutée.

Si la taille de la requête est considérée comme très grande, cela signifie que la requête nécessitera trop de temps et de ressources. Dans ce cas, le comportement du bouton *Continuer* dépend de la configuration définie par l'administrateur DBMS (celui-ci peut par exemple refuser les requêtes considérées comme trop volumineuses). Dans les deux cas, il est recommandé de consulter l'administrateur.

Le mode Explain n'a pas été configuré

Si votre administrateur DBMS n'a pas configuré le mode Explain, une fenêtre contextuelle s'ouvre lorsque vous essayez d'accéder aux données.

Vous devez contacter votre administrateur qui vous informera de la méthode à suivre et du mode à configurer.

S'il valide l'exécution de la requête, vous souhaiterez peut-être que toutes les requêtes de même durée soient exécutées sans validation. Dans ce cas, cochez la case *Ne plus demander de validation pour ce type de requête*. Ainsi, le message de validation s'affichera uniquement pour les requêtes volumineuses. Cette configuration

sera uniquement utilisée pour la session en cours. Une fois l'application fermée, elle sera perdue. Si vous souhaitez que la configuration soit permanente, contactez votre administrateur DBMS.

5 Automated Analytics Modeler

5.1 Ouverture ou suppression d'un modèle

5.1.1 Ouverture ou suppression d'un modèle

Utilisez le panneau *Ouverture d'un modèle* pour gérer des modèles que vous avez auparavant créé avec Automated Analytics ou pour exposer des modèles générés par Predictive Factory.

> **i** Remarque
>
> Par défaut, les modèles générés par Predictive Factory ne s'affichent pas. Si vous souhaitez les voir s'afficher dans la liste, cochez *Afficher les modèles générés par Predictive Factory*. Pour les différencier des modèles créées avec Automated Analytics, ils apparaîtront grisés.

Pour charger un modèle

1. Dans la liste *Type de données*, sélectionnez le type de stockage du modèle.
2. Utilisez le bouton *Parcourir* se trouvant à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où est stocké le modèle. Si la base de données est protégée, saisissez les identifiants corrects dans les champs *Utilisateur* et *Mot de passe* de la fenêtre *Sélection de la source de données*. Tous les modèles stockés dans la source sélectionnée seront affichés. La table suivante présente les informations fournies pour chaque modèle.

Colonne	Description	Valeurs
Nom	Nom sous lequel le modèle a été enregistré	Chaîne de caractères

Colonne	Description	Valeurs
Classe	Classe du modèle étant le type de modèle	• *Kxen.Classification* : classification/régression avec cible nominale • *Kxen.Regression* : classification/régression avec cible continue • *Kxen.Segmentation* : segmentation avec mode SQL • *Kxen.Clustering* : Segmentation sans mode SQL • *Kxen.TimeSeries* : série chronologique • *Kxen.AssociationRules* : règles d'association • *Kxen.SimpleModel* : Modèles multi-cible de classification/régression et de segmentation, tout autre modèle
Version	Nombre de version du modèle lorsque le modèle a été enregistré plusieurs fois	Nombre entier commençant à 1
Date	Date d'enregistrement du modèle	Date et heure au format aaaa-mm-jj hh:mm:ss
Commentaire	Commentaire défini par un utilisateur facultatif défini pouvant être utilisé pour identifier le modèle	Chaîne de caractères

> **i** Remarque
>
> Lors de l'affichage du contenu de la base de données, cliquez sur le bouton *Actualiser* pour voir tous les modèles, en particulier ceux qui ont été créés récemment.

3. Sélectionnez le modèle à charger.
4. Cliquez sur le bouton *Ouvrir*. L'écran *Utiliser le modèle* s'affiche. Toutes les options d'affichage des résultats du modèle seront disponibles.
   Plusieurs tables ou fichiers doivent pouvoir ouvrir un modèle. Si certains sont manquants, un message s'affiche, indiquant les noms des éléments manquants. Cela peut arriver lorsque vous envoyez le modèle pour le partager et que vous ne joignez pas toutes les tables ou les fichiers supplémentaires. La table suivante répertorie les éléments essentiels à l'ouverture d'un modèle.

Nom du fichier	Description	Utilisé par
KxAdmin	répertorie tous les modèles contenus dans le dossier/la base de données avec les informations supplémentaires (date, version, nom du modèle, commentaires)	tous les modèles créés avec Automated Analytics

Nom du fichier	Description	Utilisé par
<Model_name>	fichier nommé selon le modèle et contenant toutes les données du modèle sauf les informations des graphiques. Les graphiques sont stockés dans les tables supplémentaires (voir ci-dessous)	tous les modèles créés avec Automated Analytics
KxInfos	indique quels sont les tables supplémentaires requises par le modèle	tous les modèles créés avec Automated Analytics
KxOlapCube	stocke le cube OLAP utilisé par l'arbre de décision lorsque l'option Classification/Régression en tant que Décision est activé	Modèles de classification/régression avec arbre de décision
KxLinks	contient les liens des graphiques du modèle	Modèles Social uniquement
KxNodes	répertorie tous les nœuds de tous les graphiques et leurs dimensions	Modèles Social uniquement
KxCommunities	correspond aux nœuds vers leurs communautés, si la détection de la communauté a été activée	Modèles Social uniquement

Pour supprimer un modèle

1. Sélectionnez les modèles que vous souhaitez supprimer (utilisez la touche CTRL pour sélectionner plusieurs modèles).
2. Cliquez sur le bouton *Supprimer les éléments sélectionnés*.

> ⚠ Attention
>
> Pour éviter des erreurs, les modèles créés avec Predictive Factory ne peuvent pas être supprimés.

5.1.2 Ouverture ou suppression d'une manipulation de données

Ce panneau permet de recharger une manipulation de données existante pour la modifier ou simplement visualiser ses paramètres.

- Pour recharger un jeu de données d'analyse
  1. Utilisez le bouton *Parcourir* correspondant au champ *Source de base de données* pour sélectionner la base de données contenant la manipulation de données que vous souhaitez ouvrir. Si la base de données est protégée, saisissez les informations de connexion dans les champs *Utilisateur* et *Mot de passe* de la fenêtre *Sélection de la source de données*.
     Les manipulations de données stockées dans la base de données sélectionnée s'affichent.

> **i Remarque**
>
> Cliquez sur le bouton *Actualiser* pour voir tous les modèles, en particulier ceux qui ont été créés récemment.

2. Sélectionnez la manipulation de données que vous souhaitez ouvrir.
3. Cliquez sur le bouton *Ouvrir*.

- Pour supprimer une manipulation de données existante
  1. Sélectionnez la manipulation de données que vous souhaitez supprimer (utilisez la touche CTRL pour sélectionner plusieurs manipulations de données).
  2. Cliquez sur le bouton *Supprimer les éléments sélectionnés*.

5.1.3 Ouverture d'un modèle Social

Ce panneau vous permet de recharger des modèles enregistrés antérieurement.

- Pour ouvrir un modèle
  1. Dans la liste *Type de données*, sélectionnez le type de stockage du modèle.
  2. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où le modèle est enregistré. Dans le cas d'une base de données protégée, saisissez les informations d'identification dans les champs *Utilisateur* et *Mot de passe* affichés dans la fenêtre *Sélection d'une source de données*. Les modèles enregistrés dans cette source seront affichés. Le tableau ci-dessous détaille l'ensemble des informations fournies pour chaque modèle.

Colonne	Description	Valeurs
Nom	Nom sous lequel le modèle a été enregistré	Chaîne de caractères
Type	Type du modèle	• Kxen.SimpleModel : modèles de classification, régression ou segmentation multi-cibles, modèles d'analyse de réseaux sociaux antérieurs à la version 6.0 ou générés à partir de scripts et tout autre modèle • Kxen.Social : modèle d'analyse de réseaux sociaux
Version	Numéro de version du modèle lorsque celui-ci a été enregistré plusieurs fois	Entier commençant à 1
Date	Date d'enregistrement du modèle	Date et heure au format aaaa-mm-jj hh:mm:ss
Commentaire	Commentaire facultatif saisi par l'utilisateur pour faciliter l'identification du modèle	Chaîne de caractères

> **i Remarque**
>
> Lorsque vous affichez le contenu d'une base de données, vous pouvez être amené à cliquer sur le bouton *Actualiser* pour voir tous les modèles, en particulier les plus récents.

3. Sélectionnez le modèle à ouvrir.
4. Cliquez sur le bouton *Ouvrir*. L'écran *Exécution* s'affiche. Toutes les options pour afficher les résultats sont disponibles.
   Plusieurs tables ou fichiers sont nécessaires pour pouvoir ouvrir un modèle. Si certains manquent, un message s'affiche indiquant leurs noms. Cela peut se produire si vous envoyez le modèle pour le partager sans joindre les tables ou les fichiers supplémentaires. Le tableau ci-dessous liste les éléments indispensables pour pouvoir ouvrir un modèle.

Nom du fichier	Description
KxAdmin	liste tous les modèles contenus dans le dossier ou la base de données ainsi que leurs informations de base (date, version, nom du modèle, commentaires)
<Nom_modèle>	fichier nommé d'après le modèle et contenant toutes les données à l'exception des informations des graphes. Ces dernières sont stockées dans des tables ou fichiers supplémentaires (voir ci-dessous)
KxInfos	indique quelles tables additionnelles sont utilisées par le modèle
KxOlapCube	stocke le cube OLAP utilisé par l'arbre de décision lorsque l'option Classification/Régression en tant que Décision est activé dans Automated Analytics Modeler
KxLinks	contient les liens des graphes du modèle
KxNodes	liste l'ensemble des nœuds de tous les graphes et leurs attributs
KxCommunities	contient les correspondances entre les nœuds et leur communauté lorsque la détection des communautés a été activée

- Pour supprimer un modèle
  1. Sélectionnez le modèle que vous voulez supprimer (utilisez la touche CTRL du clavier pour sélectionner plusieurs modèles).
  2. Cliquez sur le bouton *Supprimer*.

5.2 Ajouter une fonctionnalité de modélisation

Lorsque vous créez un modèle, vous pouvez soit simplement créer des agrégations soit ajouter des transformations supplémentaires telles qu'une classification/régression ou une segmentation.

La liste ci-dessous présente les options de ce panneau :

- *Ajouter une classification/régression* crée une table d'agrégation et construit un modèle de classification/régression dedans.
- *Ajouter une segmentation ou un regroupement*, qui permet de créer une table d'agrégation puis de créer un modèle K2S à partir de cette table.
- *Pas d'ajout de calcul* crée une table d'agrégation. La création d'une agrégation sans aucune modélisation vous permet d'enregistrer la table d'agrégation pour pouvoir l'utiliser comme données d'entrée pour construire plusieurs modèles sur les mêmes agrégations.

5.3 Sélection des données

5.3.1 Sources de données prises en charge

Automated Analytics prend en charge les sources de données suivantes :

- Les fichiers texte (ou fichiers plats) dans lesquels les données sont séparées par un délimiteur, tel qu'une virgule dans des fichiers `.csv` (Comma Separated Value).

  > **! Restriction**
  >
  > Lors de l'accès aux données contenues dans les fichiers `.csv`, Automated Analytics prend uniquement en charge CR + LF (très répandu sous Microsoft Windows) ou LF (très répandu sous Linux) pour les sauts de ligne.

- Les systèmes de gestion de base de données sont accessibles via ODBC.

  > **i Remarque**
  >
  > Pour afficher une liste des sources compatibles ODBC, voir le document SAP Product Availability Matrix (PAM) à l'adresse http://service.sap.com/sap/support/pam.
  >
  > Pour en savoir plus sur l'utilisation de SAP HANA, voir les informations associées ci-dessous.
  >
  > Pour configurer les outils de modélisation Automated Analytics pour accéder aux données de votre système de gestion de base de données, référez-vous aux guides *Connecting your Database Management System on Windows* ou *Connecting your Database Management System on Linux*.

- Fichiers SAS

5.3.1.1 SAP HANA en tant que source de données

Vous pouvez utiliser les bases de données SAP HANA en tant que sources de données dans Data Manager et pour tous les types d'analyses de modélisation dans Modeler : Classification/Régression, Mise en cluster, Séries temporelles, Règles d'association, Social et Recommendation.

Les tables SAP HANA ou les vues SQL	Dans le nœud *Catalogue* de la base de données SAP HANA.
Tous types de vues SAP HANA	Dans le nœud *Contenu* de la base de données SAP HANA. Une vue SAP HANA est un regroupement virtuel prédéfini de colonnes de table qui permet d'accéder aux données correspondant à un besoin de gestion spécifique. Les vues sont spécifiques au type de tables incluses et au type de calculs appliqués aux colonnes. Par exemple, une vue analytique est conçue sur une table de faits et des vues d'attributs associées. Une vue de calcul exécute une fonction sur des colonnes lors de l'accès à la vue. **! Restriction** • Les vues analytiques et de calcul qui utilisent la fonction de mappage des variables (disponible à partir de SAP HANA SPS 09) ne sont pas prises en charge. • Il est impossible de modifier les données dans des vues SAP HANA à l'aide de Automated Analytics.

Tables virtuelles Smart Data Access	Grâce à Smart Data Access, vous pouvez exposer des données à partir de tables source distantes en tant que tables virtuelles et les combiner avec les tables normales HANA. Cela vous permet d'accéder aux sources de données qui ne sont pas, à l'origine, prises en charge par l'application ou de combiner des données provenant de plusieurs sources hétérogènes. **⚠ Attention** Afin d'utiliser les tables virtuelles en tant que jeux de données d'entrée pour entraîner ou appliquer un modèle d'une part, ou comme jeux de données de sortie pour appliquer un modèle d'autre part, vous devez vérifier que les conditions suivantes sont remplies : • Le mode Application in-database n'est pas utilisé. • La table de destination pour le stockage des valeurs prédites existe dans la source distante avant d'appliquer le modèle. • La structure de la table distante, c'est-à-dire les noms et types de colonnes, doit correspondre exactement à ce qui est attendu par rapport aux options de génération. Si ce n'est pas le cas, une erreur sera générée. **⚠ Attention** Dans Data Manager, utilisez les tables virtuelles avec prudence car les requêtes générées peuvent être complexes. Smart Data Access ne sera peut-être pas en mesure de déléguer une large part du traitement à la source sous-jacente en fonction des fonctionnalités de la source. Cela peut avoir un impact sur la performance.

Conditions prérequises

Vous devez connaître le nom de la source ODBC ainsi que les informations de connexion à votre base de données SAP HANA. Pour en savoir plus, contactez votre administrateur SAP HANA.

Outre les autorisations requises pour effectuer des requêtes sur une vue SAP HANA, vous devez disposer de l'autorisation `SELECT` sur le schéma `_SYS_BI`, qui contient les métadonnées sur les vues. Pour en savoir plus sur les notions de sécurité, voir les guides SAP HANA.

5.3.2 Données à modéliser

Le jeu de données à modéliser est spécifié dans cet écran. Les données d'origine peuvent être au format texte délimité par des tabulations ou des virgules, extraites d'une source ODBC ou encore être dans un format exclusif (SPSS, Excel, etc.).

- Utilisation d'un fichier ou d'une table de base de données [page 87]
- Sélection d'une vue SAP HANA [page 88]
- Sélection d'une table SAP HANA ou d'une vue SQL en tant que source de données [page 89]
- Utilisation de Data Manager [page 89]
- Stratégie de partition [page 90]
- Création d'une variable cible [page 90]
- Métadonnées [page 90]

5.3.2.1 Utilisation d'un fichier ou d'une table de base de données

Pour utiliser un fichier ou une table de base de données :

1. Cochez l'option *Utiliser un fichier ou une table de base de données*.
2. Dans la liste *Type de données*, sélectionnez le type de données auquel vous souhaitez accéder. L'accès à certains types de données nécessite une licence spécifique.
3. Utilisez le bouton *Parcourir* correspondant au champ *Dossier* pour sélectionner le dossier ou la base de données contenant les données désirées. En cas de base de données protégée, vous devrez saisir le nom d'utilisateur et le mot de passe lors de la sélection de la base de données.
   Lorsque le format des données le permet, vous disposez des options suivantes dans la fenêtre *Sélection de données* :
   - : aller au répertoire parent
   - : créer un dossier
   - : supprimer le dossier sélectionné (cela n'est possible que pour des dossiers vides)
4. Utilisez le bouton *Parcourir* correspondant au champ *Jeux de données* pour sélectionner le jeu de données à utiliser. Les jeux de données d'analyse créées avec Automated Analytics sont affichés en haut de la liste et indiqués par l'icône .
5. Si le jeu de données ne contient pas de variable cible, vous pouvez la créer en cliquant sur le bouton *Créer une cible*.
6. Le bouton Loupe permet de visualiser les données et de calculer les statistiques du jeu de données sélectionné.
7. Le bouton *Paramètres avancés* vous permet de définir certains paramètres avancés tels que l'échantillonnage ou les paramètres de fichiers.
8. Cochez l'option *Mode fabrique de données*. Cette option peut être utile lorsque vous utilisez des métadonnées différentes selon les équipes avec lesquelles vous travaillez.

5.3.2.1.1 Sélection d'une vue SAP HANA

Vous pouvez choisir une vue analytique ou de calcul SAP HANA en tant que source de données pour votre modèle.

1. Dans l'écran *Données à modéliser*, sélectionnez *Base de données* dans la liste *Type de données*.
2. Cliquez sur le bouton *Parcourir* à gauche du champ *Dossier*, puis sélectionnez la base de données SAP HANA que vous souhaitez utiliser dans la liste *Sélectionnez une source pour les données*.
3. S'il s'agit de la première connexion à cette base de données, saisissez les informations de connexion dans les champs *Utilisateur* et *Mot de passe* , puis cliquez sur *Se connecter*.

   Le contenu de la base de données s'affiche dans la partie gauche de la fenêtre.
4. Ouvrez le nœud *Contenu*. Utilisez l'arborescence pour parcourir les packages et sélectionnez celui qui contient la vue SAP HANA que vous souhaitez utiliser en tant que source de données pour votre modèle, puis cliquez sur *OK*.
5. Cliquez sur le bouton *Parcourir* à gauche du champ *Jeu de données*.
6. Dans la boîte de dialogue *Sélection de données*, sélectionnez la vue que vous souhaitez utiliser.

   Il est possible de filtrer la liste de vues en saisissant une partie du nom de la vue dans le champ situé sous la liste.
7. Cliquez sur *OK*.

   Si la vue SAP HANA que vous avez sélectionnée requiert la définition de certaines valeurs, la fenêtre *Valeurs utilisateur* s'ouvre lorsque vous cliquez sur *Afficher les données* ou *Suivant* sur l'écran *Données à modéliser*. Les variables pour lesquelles vous devez saisir une valeur sont répertoriées à gauche du panneau. Le tableau suivant présente en détails la méthode d'identification des différentes variables dans la liste. Dans cette rubrique, le terme variable est utilisé à la fois pour les variables et les paramètres d'entrée.

Type de variable	Noté comme	Signalé par
Variable obligatoire	obligatoire	Une croix rouge
Variable facultative	facultative	
Variable avec une valeur par défaut	valeur par défaut	Une coche verte
Variable avec une valeur définie par l'utilisateur		Une coche verte

   Cochez la case *Afficher uniquement obligatoires* pour afficher uniquement les variables obligatoires dans la liste.
8. Dans la liste, sélectionnez les variables pour lesquelles vous souhaitez définir les valeurs.

   Pour cela, le champ approprié est affiché à droite du panneau. Selon le type de valeur requis, il est possible que plusieurs types de champs s'affichent. Par exemple, si une valeur unique est requise, une liste des opérateurs et un champ de saisie de valeur s'affichent. Si une plage de valeurs est requise, vous verrez apparaître deux champs pour les limites de plage et une liste des opérateurs.
9. Certaines variables contiennent une liste de valeurs possibles prédéfinies. Dans ce cas, la liste des valeurs pour la variable actuelle est affichée dans la partie droite du panneau. Sélectionnez la valeur que vous souhaitez définir pour cette variable ; elle sera affichée dans le champ de valeur.
10. Utilisez le bouton *Restaurer* pour revenir à la valeur par défaut de la variable en cours.

11. Une fois toutes les valeurs requises saisies, cliquez sur *OK*.

    Après avoir accédé à la vue SAP HANA, les valeurs définies seront mémorisées et utilisées par défaut ultérieurement lors du processus de modélisation.

5.3.2.1.2 Sélection d'une table SAP HANA ou d'une vue SQL en tant que source de données

1. Dans l'écran *Données à modéliser*, sélectionnez *Base de données* dans la liste *Type de données*.
2. Cliquez sur le bouton *Parcourir* à gauche du champ *Dossier*.
3. Dans la boîte de dialogue *Sélection de données*, sélectionnez la base de données SAP HANA que vous souhaitez utiliser dans la liste *Sélectionnez une source pour les données*.
4. S'il s'agit de la première connexion à cette base de données, saisissez les informations de connexion dans les champs *Utilisateur* et *Mot de passe* , puis cliquez sur *Se connecter*.

   Le contenu de la base de données s'affiche dans la partie gauche de la fenêtre.
5. Ouvrez le nœud *Catalogue*, sélectionnez le schéma contenant la table ou la vue SQL que vous souhaitez utiliser en tant que source de données pour votre modèle, puis cliquez sur *OK*.
6. Cliquez sur le bouton *Parcourir* à gauche du champ *Jeu de données*.
7. Dans la boîte de dialogue *Sélection de données*, sélectionnez la table ou la vue SQL que vous souhaitez utiliser.

   Vous pouvez filtrer la liste de tables en saisissant une partie du nom de table ou de vue SQL dans le champ situé sous la liste.
8. Cliquez sur *OK*.

5.3.2.2 Utilisation de Data Manager

À l'aide de Data Manager, vous pouvez sélectionner un enregistrement analytique et une population estampillée comme source de données et créer automatiquement le jeu de données d'analyse correspondant.

- Pour sélectionner une source de données avec Data Manager
  1. Cochez la case *Utiliser Data Manager*.
  2. Utilisez le bouton *Parcourir* du champ *Source de base de données* pour sélectionner la base de données contenant les données. Si la base de données est protégée, saisissez le nom d'utilisateur et le mot de passe dans la fenêtre *Sélection d'une source de données* avant de valider votre sélection.
  3. Utilisez le bouton *Parcourir* correspondant au champ *Enregistrement analytique* pour sélectionner l'enregistrement analytique à utiliser. Dans la fenêtre *Sélection d'une source de données*, les enregistrements analytiques disponibles sont affichés en tête de liste sous l'élément *Enregistrement analytique*.
  4. Utilisez le bouton *Parcourir* correspondant au champ *Population estampillée* pour sélectionner la population estampillée à utiliser. Dans la fenêtre *Sélection d'une source de données*, les populations estampillées disponibles sont affichées en tête de liste sous l'élément *Population estampillée*.
  5. Si l'enregistrement analytique ne contient pas de variable cible, vous pouvez la créer en cliquant sur le bouton *Créer une cible*.

6. Le bouton Loupe permet de visualiser les données et de calculer les statistiques du jeu de données sélectionné.

5.3.2.3 Stratégie de partition

Deux jeux de données obligatoires et un facultatif sont nécessaires pour la modélisation :

- *Entraînement* : utilisé pour l'entraînement des modèles.
- *Validation* : utilisé pour garantir la cohérence (robustesse) des modèles.
- *Test* (facultatif) : utilisé pour calculer certains estimateurs de qualité du modèle.

La stratégie de partition *Aléatoire sans test* est utilisée par défaut, mais vous pouvez choisir d'appliquer une autre stratégie en cliquant sur le bouton *Stratégie de partition*.

5.3.2.4 Création d'une variable cible

Si votre jeu de données ne contient pas de variable cible répondant à votre problème, vous pouvez en créer une à la volée à partir des variables présentes dans votre base de données. Cette nouvelle variable peut être créée comme :

- une fonction
- un agrégat
- une condition
- une normalisation
- une table de recherche
- ou une expression libre.

- Pour créer une variable cible
  - Cliquez sur la flèche à droite du bouton *Créer une cible* pour choisir le type d'opération à utiliser pour créer la nouvelle variable. Par défaut, le bouton *Créer une cible* affiche l'éditeur d'expressions, vous permettant ainsi de saisir une expression libre.

5.3.2.5 Métadonnées

Cochez l'option *Mode fabrique de données*. Cette option peut être utile lorsque vous utilisez des métadonnées différentes selon les équipes avec lesquelles vous travaillez.

Vous pouvez également modifier ce paramètre dans les préférences d'Automated Analytics, accessibles par le bouton situé en bas à gauche de chaque panneau dans le *Référentiel de métadonnées*.

5.3.3 Description des données

Pour que vos données soient interprétables et analysables par les fonctions SAP Predictive Analytics, elles doivent être décrites. En d'autres termes, le fichier de description doit spécifier la nature de chaque variable en déterminant leur format de stockage et leur type.

- Définitions [page 91]
- Description des données [page 94]
- Enregistrement d'une description [page 93]
- Aperçu des données [page 95]
- Modification des données [page 96]
- Structure [page 99]
- Ajout d'un filtre au jeu de données [page 105]
- Codage des variables [page 106]

5.3.3.1 Définitions

Description d'une variable

La description précise d'une variable permet aux fonctionnalités d'Automated Analytics d'effectuer les calculs appropriés lors de la génération d'un modèle en vue d'obtenir les meilleurs résultats.

Chaque variable ou attribut est décrit par les informations suivantes :

Champ	Description	Valeurs
Nom	nom de la variable (ne peut être modifié)	

Champ	Description	Valeurs
Type	type de valeurs enregistrées dans cette variable Remarque : Lorsque vous déclarez une variable comme date ou date et heure, codeur de date extrait automatiquement de cette variable les informations de date telles que le jour du mois, l'année, le trimestre, etc. Des variables supplémentaires contenant ces informations sont créées lors de la génération et utilisées comme variables d'entrée du modèle. Le codeur de date est désactivé pour les modèles de série temporelle.	*Nombre* : nombres "calculables" (attention : les numéros de téléphone ou de compte ne doivent pas être considérés comme des nombres). *Entier* : nombres entiers *Chaîne* : chaînes de caractères *Date et heure* : date et heure *Date* : dates
Valeur	décrit le type de valeur **i Remarque** Lorsque vous créez un modèle d'encodage textuel, il faut que le jeu de données contienne au moins une variable textuelle pour pouvoir passer au panneau suivant.	*Continue* : variable numérique à partir de laquelle une moyenne, une variance, etc. peuvent être calculées *Nominale* : variable de catégorie (seule valeur possible pour une chaîne) *Ordinale* : variable numérique discrète dont l'ordre relatif est important *Textuelle* : variable textuelle contenant des expressions, des phrases ou des textes complets
Clé	variable clé ou identifiant pour l'enregistrement Il est possible de déclarer des clés multiples qui seront construites selon l'ordre indiqué (1-2-3-...).	0 : la variable n'est pas un identifiant, 1 : identifiant primaire, 2 : identifiant secondaire, ...
Ordre	indique si cette variable représente un ordre naturel. Si la valeur est 1, la variable est utilisée dans les expressions SQL dans une condition "ORDER BY". **i Remarque** définir une variable comme variable d'ordre aura une incidence sur le choix d'une stratégie de partition. Seules les stratégies séquentielles peuvent être utilisées avec une variable d'ordre.	0 : la variable ne représente pas un ordre naturel 1 : la variable représente un ordre naturel

Champ	Description	Valeurs
Inconnu	Chaîne utilisée dans les données pour représenter une valeur manquante	Par exemple : 999 ou #Vide
Groupe	Indique le nom du groupe auquel appartient cette variable. Vous pouvez affecter une variable à seulement un groupe. Toutes les variables d'un groupe contiennent la même information. Différents groupes ne peuvent donc pas être croisés pour les modèles d'un degré supérieur à 1.	
Description	étiquette de description supplémentaire pour la variable	

Bibliothèque de variables

La bibliothèque de variables vous permet de stocker les descriptions de variables que vous avez déjà utilisées. Vous pourrez ainsi les réutiliser automatiquement lors de la description d'un jeu de données par analyse.

5.3.3.2 Enregistrement d'une description

Vous pouvez enregistrer la description des variables soit dans la bibliothèque de variables, soit dans un fichier ou une table de base de données.

Dans la bibliothèque de variables

Vous pouvez choisir d'enregistrer les descriptions de toutes les variables, ou sélectionner uniquement celles qui vous intéressent. Vous pouvez aussi supprimer des variables de la bibliothèque de variables.

Si vous avez défini un référentiel de métadonnées unique (option accessible via le bouton *Aide*, puis *Modifier les options...*, onglet *Référentiel de métadonnées*), les descriptions sont enregistrées dans ce référentiel sinon elles sont enregistrées au même emplacement que le jeu de données que vous décrivez.

- Pour enregistrer toutes les descriptions :
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Enregistrer dans la bibliothèque de variables*.
  2. Sélectionnez l'option *Toutes les variables*.
- Pour enregistrer uniquement certaines descriptions :
  1. Sélectionnez la variable dont vous voulez enregistrer la description. Pour sélectionner plusieurs variables, utilisez la touche `Ctrl`.
  2. Cliquez sur le bouton *Enregistrer dans la bibliothèque de variables*.
  3. Sélectionnez l'option *Variables sélectionnées*.
- Pour supprimer de la bibliothèque de variables :

- Sélectionnez la variable à supprimer de la bibliothèque de variables et cliquez sur Supprimer de la bibliothèque de variables.

Dans un fichier ou une table de base de données

Vous pouvez enregistrer l'ensemble des descriptions du jeu de données dans un fichier ou dans une table de base de données.

- Pour enregistrer la description du jeu de données dans un fichier ou une table :
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Enregistrer la description*. La fenêtre *Enregistrer une description de...* s'affiche.
  2. Dans la liste *Type de données*, sélectionnez le format dans lequel vous voulez enregistrer la description.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour choisir le dossier ou la base de données où vous souhaitez enregistrer la description.
  4. Dans le champ *Description*, saisissez le nom du fichier ou de la table où vous souhaitez enregistrer la description.
  5. Cliquez sur *OK*.

5.3.3.3 Description des données

Il existe deux façons de décrire les données :

- en analysant le jeu de données ;
- en se servant d'une description du jeu de données complet préalablement enregistré sous forme de fichier ou de table de base de données.

Analyse du jeu de données

L'analyse du jeu de données d'entraînement vous permettra de commencer avec une description obtenue soit par l'analyse des cents premières lignes de données, soit par l'utilisation d'une description issue de la bibliothèque de variables si une variable de même nom s'y trouve. Vous pouvez ensuite modifier les descriptions proposées (en vérifiant surtout le genre des variables).

> **i** Remarque
>
> Pour les fichiers texte, il est bon d'avoir une première ligne d'en-tête contenant les noms des variables.

Lorsque vous créez la description par *analyse* automatique, vérifiez chaque variable afin de valider la description comme indiqué dans la définition Description d'une variable.

- Pour analyser le jeu de données
  - Cliquez sur le bouton *Analyser* situé dans l'onglet *Accueil* du ruban.

Les variables dont la description provient de la bibliothèque de variables sont signalées par l'icône dans la colonne *Index*.

Chargement d'une description du jeu de données

Pour décrire le jeu de données, vous pouvez utiliser un fichier ou une table de base de données défini(e) précédemment et contenant les descriptions de toutes les variables.

- Pour ouvrir une description de jeu de données
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Ouvrir une description*.
  2. Dans la fenêtre *Ouvrir une description pour...*, sélectionnez le format de la description dans la liste déroulante *Type de données*.
  3. Utilisez le bouton *Parcourir* à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où doit être stockée la description.
  4. Utilisez le bouton *Parcourir* situé à droite du champ *Description* pour sélectionner le fichier ou la table contenant la description.
  5. Cliquez sur *OK*. La fenêtre se ferme et la description s'affiche dans le panneau principal. Vous pouvez alors la modifier si nécessaire.

5.3.3.4 Aperçu des données

L'écran *Description de données* vous permet de consulter et d'afficher des informations sur le jeu de données ou les variables.

Le tableau suivant liste les options disponibles :

L'option...	affiche...	Instructions
Aperçu des données	Contenu de la manipulation des données. Il vous permet de vérifier si les résultats correspondent à votre attente. Dans ce panneau, vous pouvez trier les données par colonne et sélectionner les lignes à afficher.	• Cliquez sur le bouton Aperçu des données.
Propriétés	Propriétés de la variable que vous avez sélectionnée, telles que le nom de la variable, le stockage, le type de valeur de la variable, la clé, l'ordre, les valeurs manquantes, le nom du groupe auquel la variable appartient et une description supplémentaire.	• Cliquez sur le bouton Propriétés.

5.3.3.5 Modification des données

Dans l'écran *Description des données*, vous pouvez accéder à différentes options vous permettant de :

- spécifier la nature de chaque variable en déterminant leur format de stockage et leur type ;
- modifier ou remettre à zéro la valeur manquante ;
- traduire les catégories dans la langue de votre choix ;
- modifier les variables composites ou les groupes.

Modifier les variables

- Pour modifier le stockage d'une variable...
  - Avec l'onglet *Édition* du ruban :
    1. Sélectionnez la variable dont vous souhaitez modifier le stockage.
    2. Dans l'onglet *Édition* du ruban, cliquez sur *Stockage*.
    3. Sélectionnez le stockage dans la liste.
  - Avec la colonne *Stockage* de la description :
    1. Sélectionnez la variable dont vous souhaitez modifier le stockage.
    2. Dans le champ *Stockage*, sélectionnez le stockage dans la liste.
  - avec l'utilisation du menu contextuel :
    1. Faites un clic droit sur la variable dont vous souhaitez modifier le stockage. Un menu contextuel apparaît.
    2. Cliquez sur *Stockage* et sélectionnez le stockage dans la liste.
- Pour modifier le type d'une variable...
  - Avec l'onglet *Édition* du ruban :
    1. Sélectionnez la variable dont vous souhaitez modifier le type.
    2. Dans l'onglet *Édition* du ruban, cliquez sur *Valeur*.
    3. Sélectionnez le type dans la liste.
  - Avec la colonne *Valeur* de la description :
    1. Sélectionnez la variable dont vous souhaitez modifier le type.
    2. Dans le champ *Valeur*, sélectionnez le type dans la liste.
  - avec l'utilisation du menu contextuel :
    1. Faites un clic droit sur la variable dont vous souhaitez modifier le type. Un menu contextuel apparaît.
    2. Cliquez sur *Type* et sélectionnez le type dans la liste.
- Pour modifier la valeur manquante...
  - Avec l'onglet *Édition* du ruban :
    1. Sélectionnez la variable dont vous souhaitez modifier la valeur manquante.
    2. Dans l'onglet *Édition* du ruban, cliquez sur *Modifier la valeur manquante*. Le panneau *Modifier la valeur manquante* s'affiche.
    3. Tapez la valeur manquante.
    4. Cliquez sur *OK*.

- Avec la colonne *Inconnu* de la description :
  1. Sélectionnez la variable dont vous souhaitez modifier la valeur manquante.
  2. Dans le champ *Inconnu*, tapez la valeur manquante.
- avec l'utilisation du menu contextuel :
  1. Faites un clic droit sur la variable dont vous souhaitez modifier la valeur manquante. Un menu contextuel apparaît.
  2. Cliquez sur *Valeurs manquantes* et sélectionnez *Modifier la valeur manquante* dans la liste. Le panneau *Modifier la valeur manquante* s'affiche.
  3. Tapez la valeur manquante.
  4. Cliquez sur *OK*.

- Pour remettre la valeur manquante par défaut...
  - Avec l'onglet *Édition* du ruban :
    1. Sélectionnez la variable dont vous souhaitez modifier la valeur manquante.
    2. Dans l'onglet Édition du ruban, cliquez sur *Valeur manquante par défaut*.
  - Avec la colonne *Inconnu* de la description :
    1. Sélectionnez la variable dont vous souhaitez modifier la valeur manquante.
    2. Dans le champ *Inconnu*, effacez la valeur manquante.
  - avec l'utilisation du menu contextuel :
    1. Faites un clic droit sur la variable dont vous souhaitez modifier la valeur manquante. Un menu contextuel apparaît.
    2. Cliquez sur *Valeurs manquantes* et sélectionnez *Réinitialiser la valeur manquante* dans la liste.

Traduction des catégories

Vous pouvez traduire les catégories des variables nominales, enregistrer la traduction ou charger une traduction existante. Cette traduction n'influence pas la structure de la variable, qui doit être définie en fonction des valeurs originales de la variable.

> **i** Remarque
>
> La variable "Clé cible", utilisée par exemple dans les paramètres avancés, ne prend pas en compte une éventuelle traduction quand les valeurs possibles de cette variable sont affichées. Pour cette raison des valeurs entrées manuellement ne peuvent pas être traitées correctement, si elles ne correspondent pas aux valeurs d'origine.

- Traduire les catégories de variables
  1. Sélectionnez la variable nominale à traduire. Vous pouvez :
     - Soit aller sur l'onglet Édition et cliquer sur *Traduire les catégories*.
     - Soit effectuer un clic droit pour afficher un menu contextuel, puis cliquer sur *Traduire les catégories*.
       Le panneau *Traduire les catégories de <nom_de_la_variable>* s'affiche.
  2. Cliquez sur le bouton *Ajouter une langue* pour sélectionner les langues que vous voulez traduire. Les langues ajoutées s'affichent en colonnes.

     > **i** Remarque
     >
     > Par défaut, la langue de l'interface utilisateur est affichée sous forme de colonne.

3. Cliquez sur le bouton pour extraire les catégories de variables du jeu de données.
4. Traduisez les catégories.

> **i Remarque**
>
> - Vous n'avez pas besoin de renseigner tous les champs.
> - Vous pouvez ajouter ou supprimer des catégories avec les boutons *Ajouter une catégorie* et *Supprimer la catégorie*.

5. Cliquez sur le bouton *OK*.

- Pour enregistrer la traduction des catégories
  1. Traduisez les catégories de variables comme expliqué ci-dessus.
  2. Cliquez sur le bouton *Enregistrer*.
  3. Choisissez un *Type de données*.
  4. Sélectionnez un *Dossier*.
  5. Entrez un *Nom* pour le fichier ou la table.
  6. Cliquez sur le bouton *OK*.
- Pour charger un fichier de traduction existant
  1. Sélectionnez la variable nominale à traduire. Vous pouvez :
     1. Soit aller à l'onglet Édition et cliquer sur le bouton *Traduire les catégories*.
     2. Soit effectuer un clic droit pour afficher un menu contextuel. Cliquez ensuite sur *Traduire les catégories*.
        Le panneau *Traduire les catégories de <nom_de_la_variable>* s'affiche.
  2. Cliquez sur le bouton *Charger*.
  3. Sélectionnez le format de la traduction dans la liste *Type de données*.
  4. Utilisez le bouton *Parcourir* à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où doit être stockée la description.
  5. Utilisez le bouton *Parcourir* à droite du champ *Table ou fichier* pour sélectionner le fichier ou la table contenant la description.
  6. Cliquez sur le bouton *OK*.
  7. Cliquez sur le bouton *Mettre à jour* pour actualiser l'affichage des catégories.
  8. Si les colonnes ne sont pas nommées correctement, utilisez les *Paramètres avancés* pour choisir la ligne d'en-tête et actualisez à nouveau.
  9. Mettez les noms des langues en correspondance avec les langues de la traduction chargée en cliquant sur les catégories et en choisissant la langue qui correspond dans le menu contextuel.
  10. Cliquez sur le bouton *OK*.

Variables composites

L'écran Description de données vous permet de créer des variables composites et de définir des groupes.

- Pour ajouter une variable composite
  1. Sélectionnez deux variables possédant une latitude et une longitude.
  2. Accédez à l'onglet *Édition* et cliquez sur *Variables composites*.

3. Cliquez sur *Ajouter* pour entrer un nouveau nom pour la variable composite que vous voulez ajouter.
4. Dans la liste *Type*, vous pouvez choisir entre *Position*, *Croisement* et *Croisement en masse*.
   Le type Position vous permet de définir deux coordonnées à partir du même emplacement.
   Sélectionnez le type Croisement pour effectuer des croisements de variables (une par une).
   Sélectionnez le type Croisement en masse pour effectuer des croisements de variables.
5. Cliquez *OK* pour valider.

- Pour modifier une variable composite
  1. Cliquez sur *Variables Composites*.
  2. Sélectionnez celle que vous souhaitez modifier.
  3. Cliquez sur *Modifier*.
- Pour supprimer une variable composite
  1. Dans *Variables Composites*, sélectionnez la variable composite que vous souhaitez supprimer.
  2. Cliquez sur *Supprimer*.
- Pour modifier le groupe
  - Cliquez sur le bouton *Définir le groupe* pour faire le croisement d'un groupe avec un autre.

> **i** Remarque
>
> Cette option ne peut être utilisée qu'avec des croisements en masse.

- Pour renseigner des groupes à partir de domaines
  - Cliquez sur le bouton *Remplir les groupes avec les domaines* pour remplir la colonne *Groupe* par les domaines des variables.

5.3.3.6 Structure

Une fois la description des données affichée dans le panneau, l'onglet *Structures* est ajouté au ruban, ce qui vous permet de modifier la structure des données des variables.

État

La dernière colonne du tableau de la description indique l'état de la structure de chaque variable. Le tableau suivant énumère les états possibles d'une structure de variable.

Icône	État	Description
	indéfini	Le codage des données va déterminer automatiquement le regroupement de catégories en fonction de leur interaction avec la variable cible.
	non modifiable	La structure d'une variable de chaîne ordinale ne peut être modifiée.

Icône	État	Description
	définie par extraction des statistiques de la variable.	L'utilisateur doit ouvrir et valider la structure.
	définie par l'utilisateur ou importée d'un modèle existant.	

Définir une structure

Il existe plusieurs façons de définir la structure d'une variable :

- en extrayant automatiquement les catégories à partir des statistiques de la variable, puis en modifiant ou en validant la structure extraite,
- en dupliquant la structure d'une autre variable sur la variable sélectionnée,
- en important automatiquement la structure à partir d'un modèle existant contenant les mêmes données.
- en créant une structure à partir de zéro.

L'option *Activer l'optimisation des regroupements basés sur la variable cible effectués par K2C* permet au codage des données de regrouper les groupes de catégories définies dans la structure de variable s'ils apportent les mêmes informations.

> **i** Remarque
>
> La traduction des catégories d'une variable n'a pas d'influence sur sa structure qui doit être définie en fonction des valeurs initiales de la variable.

Extraire la structure d'une variable à partir des statistiques

- Pour extraire la structure d'une variable...
  - avec l'utilisation de l'onglet *Structures* du ruban :
    1. Sélectionnez les variables dont vous souhaitez extraire la structure.
    2. Cliquez sur le bouton *À partir des statistiques*. Une barre de progression est affichée au cours de l'extraction de la structure. Une fois l'extraction terminée, les icônes correspondant aux variables concernées changent, indiquant le succès de l'opération et vous permettant de les identifier clairement. Vous pouvez alors modifier la structure selon vos spécifications.
  - avec l'utilisation du menu contextuel :
    1. Sélectionnez les variables dont vous souhaitez extraire la structure. Pour sélectionner plusieurs variables, utilisez la touche CTRL.
    2. Un menu contextuel s'affiche.
    3. Sélectionnez l'option *Définir une structure* > *À partir des statistiques*. Une barre de progression est affichée au cours de l'extraction de la structure. Une fois l'extraction terminée, les icônes correspondant aux variables concernées changent, indiquant le succès de l'opération et vous permettant de les identifier clairement. Vous pouvez alors modifier la structure selon vos spécifications.

Dupliquer la structure d'une autre variable

- Pour dupliquer la structure d'une autre variable sur les variables sélectionnées...

- avec l'utilisation de l'onglet *Structures* du ruban :
  1. Sélectionnez les variables dont vous souhaitez extraire la structure.
  2. Cliquez sur *À partir d'une variable*. Un panneau s'affiche.
  3. Sélectionnez la variable dans la liste.
  4. Cliquez sur *OK*.
- avec l'utilisation du menu contextuel :
  1. Faites un clic droit sur les variables dont vous souhaitez extraire la structure. Un menu contextuel s'affiche.
  2. Cliquez sur *Définir une structure* puis sélectionnez *À partir d'une variable*. Un panneau s'affiche.
  3. Sélectionnez la variable dans la liste proposée.
  4. Cliquez sur *OK*.

> **i** Remarque
>
> Un message apparaît lorsque la structure d'une variable n'est pas définie.

Importer la structure d'une variable à partir d'un modèle existant

- Pour importer la structure d'une variable à partir d'un modèle...
  - avec l'utilisation de l'onglet *Structures* du ruban :
    1. Sélectionnez les variables dont vous souhaitez extraire la structure.
    2. Cliquez sur *À partir d'un modèle* et sélectionnez soit *Variables sélectionnées*, soit *Toutes les variables* selon les variables que vous avez sélectionnées. Le panneau *Chargement d'un modèle* s'affiche.
    3. Dans la liste *Type de données*, sélectionnez le type de stockage du modèle.
    4. Utilisez le bouton *Parcourir* situé à côté du champ *Dossier* pour sélectionner le dossier ou la base de données où est stocké le modèle.
    5. Dans la liste des modèles qui s'affiche, sélectionnez le modèle que vous voulez utiliser pour importer les structures de variables.
    6. Cliquez sur le bouton *OK*.
    7. Dans la liste *Cible du modèle chargé*, sélectionnez la cible du modèle. Les variables que vous avez sélectionnées sont affichées dans une liste avec les variables correspondantes du modèle chargé. Vous pouvez ajouter ou supprimer des variables de cette liste et visualiser la structure des variables du modèle, comme décrit ci-dessous.
    8. Une fois que toutes les variables pour lesquelles vous souhaitez importer la structure sont affichées dans la liste, cliquez sur le bouton *OK*. La fenêtre de sélection se ferme et l'icône de structure est mise à jour.
  - avec l'utilisation du menu contextuel :
    1. Faites un clic droit sur les variables dont vous souhaitez extraire la structure. Un menu contextuel s'affiche.
    2. Cliquez sur *À partir d'un modèle* et sélectionnez soit *Variables sélectionnées*, soit *Toutes les variables* selon les variables que vous avez sélectionnées. Le panneau *Chargement d'un modèle* s'affiche.
    3. Référez-vous aux instructions listées au-dessus pour continuer.
- Pour ajouter une variable à la liste
  1. Dans la liste *Variable du modèle*, sélectionnez la variable que vous voulez ajouter à la liste des variables pour lesquelles la structure sera importée du modèle.
  2. Cliquez sur le bouton *Ajouter*. La variable s'affiche dans la liste.

- Pour supprimer une variable de la liste
  1. Dans la liste des variables pour lesquelles la structure sera importée du modèle, sélectionnez la variable que vous voulez supprimer.
  2. Cliquez sur le bouton *Supprimer*. La variable est supprimée de la liste et ajoutée à la liste *Variable du modèle chargé*.
- Pour visualiser la structure d'une variable définie par le modèle chargé
  - Si la variable n'a pas encore été ajoutée à la liste des variables pour lesquelles la structure sera importée du modèle :
    1. Dans la liste *Variable du modèle*, sélectionnez la variable pour laquelle vous souhaitez importer la structure.
    2. Cliquez sur le bouton *Afficher*, la structure de la variable s'ouvre dans une nouvelle fenêtre.
  - Si la variable est déjà dans la liste des variables pour lesquelles la structure sera importée du modèle :
    1. Cliquez deux fois sur la variable dont vous souhaitez voir la structure.

Créer une structure

- Pour créer ou modifier la structure d'une variable...
  - à l'aide de l'onglet *Structures* du ruban :
  - Cliquez sur *Nouvelle Structure*. La fenêtre de modification s'ouvre.
  - Utilisation de la colonne *Structure*
  - Cliquez deux fois sur l'icône *Structure* correspondant à la variable dont vous voulez modifier ou créer la structure. La fenêtre de modification s'ouvre.
    Si vous aviez extrait la structure des statistiques de la variable ou d'un modèle, les champs de la fenêtre sont déjà préremplis.

Supprimer la structure d'une variable

- Pour supprimer la structure d'une variable...
  - à l'aide de l'onglet *Structures* du ruban :
    1. Sélectionnez les variables dont vous souhaitez supprimer la structure.
    2. Cliquez sur le bouton *Supprimer la structure*.
  - Utilisation de la colonne *Structure*
    1. Cliquez deux fois sur les variables dont vous souhaitez supprimer la structure.
    2. Cliquez sur *Définir une structure* et sélectionnez *Supprimer la structure*.

Structures par type de variable

Étant donné que la structure d'une variable dépend de son type de valeur (nominale, continue, ordinale), l'éditeur de structure diffère pour chaque type.

Structure d'une variable continue

La structure d'une variable continue est définie par plusieurs intervalles composés chacun de :

- une limite inférieure (`[ ]`) ouverte ou fermée,
- une valeur minimale (*Minimum*),
- une valeur maximale (Maximum),

- une limite supérieure (`[ ]`) ouverte ou fermée.

Tous les intervalles doivent être adjacents : il ne doit y avoir aucun trou ni chevauchement entre deux intervalles.

L'option *Ajouter manquante* vous permet d'indiquer avec quel intervalle les valeurs manquantes devront être regroupées.

L'option *Inclure les données plus petites*, vous permet d'inclure dans le premier intervalle les données plus petites que celui-ci. De la même façon l'option *Inclure les données plus grandes*, vous permet d'inclure dans le dernier intervalle les données plus grandes que celui-ci.

- Pour créer un intervalle
  1. Cliquez sur le bouton *Ajouter* pour créer un intervalle. La fenêtre de modification s'ouvre.
  2. Sélectionnez le type de la borne de début en cliquant sur le bouton *]*.
  3. Saisissez la valeur minimale de l'intervalle dans le champ de texte de gauche.
  4. Saisissez la valeur maximale de l'intervalle dans le champ de texte de droite.
  5. Sélectionnez le type de la borne de début en cliquant sur le bouton *[*.
  6. Cochez la case *Manquante* si les valeurs manquantes doivent être regroupées avec cet intervalle.
  7. Cliquez sur le bouton *Oui* pour valider votre intervalle.
- Pour séparer un intervalle
  1. Sélectionnez un intervalle.
  2. Cliquez sur le bouton *Séparer*. L'intervalle sélectionné est automatiquement séparé en deux intervalles égaux.
- Pour fusionner plusieurs intervalles
  1. Sélectionnez les intervalles à fusionner. Ceux-ci doivent être adjacents.
  2. Cliquez sur le bouton *Joindre*.
- Pour supprimer un intervalle
  1. Sélectionnez un ou plusieurs intervalles. Ceux-ci doivent être adjacents.
  2. Cliquez sur le bouton *Supprimer*. Les intervalles précédent et suivant sont étendus pour inclure les valeurs contenues dans les intervalles supprimés.

Structure d'une variable ordinale

La structure d'une variable ordinale est similaire à celle d'une variable continue, à l'exception des limites qui sont toujours fermées et ne peuvent pas être modifiées.

> **i** Remarque
>
> La structure d'une variable de chaîne ordinale ne peut être modifiée.

Structure d'une variable textuelle

La structure des variables textuelles ne peut être modifiée.

Structure d'une variable nominale

La structure d'une variable nominale est composée de plusieurs groupes rassemblant les catégories de la variable.

- Pour créer un groupe de catégories
  1. Dans la section *Structure du groupe*, saisissez dans le champ de texte le nom du groupe à créer. Utilisez la touche `Ctrl` de votre clavier pour sélectionner plusieurs catégories.

2. Cliquez sur le bouton *Ajouter un nouveau groupe*. Un groupe contenant les catégories sélectionnées est créé dans la liste *Structure du groupe*.

- Pour inclure les valeurs manquantes dans un groupe
  1. Dans la liste *Structure du groupe*, sélectionnez le groupe auquel vous voulez ajouter les valeurs manquantes.
  2. Cliquez sur le bouton *Ajouter manquante*, situé sous la liste *Édition des catégories*. La catégorie KxMissing, qui représente les valeurs manquantes, est ajoutée au groupe sélectionné et le bouton *Ajouter manquante* est désactivé. Comme toute catégorie, la catégorie KxMissing peut appartenir à un seul groupe à la fois.
- Pour créer une catégorie et l'ajouter à un groupe
  1. Dans la liste *Structure du groupe*, sélectionnez le groupe auquel vous voulez ajouter la nouvelle catégorie.
  2. Dans le champ à droite du bouton *Nouvelle catégorie*, saisissez le nom de la catégorie à ajouter.
  3. Cliquez sur le bouton *Nouvelle catégorie*. La nouvelle catégorie est créée et assignée au groupe sélectionné.
- Pour supprimer un groupe
  1. Dans la liste *Structure du groupe*, sélectionnez le groupe à supprimer.
  2. Cliquez sur le bouton *Supprimer le groupe*. Toutes les catégories appartenant à ce groupe sont rajoutées à la liste *Édition des catégories*, sauf celles qui ont été créées par l'utilisateur.
- Pour supprimer une catégorie d'un groupe
  1. Dans la liste *Structure du groupe*, sélectionnez les catégories à supprimer du groupe.
  2. Cliquez sur le bouton *Supprimer la catégorie*. Toutes les catégories assignées à ce groupe réapparaissent dans la liste *Édition des catégories*, à l'exception de celles créées par l'utilisateur.

> **i** Remarque
>
> - Pour permettre au codage des données (K2C) de réunir les groupes de catégories définis dans la structure des variables s'ils apportent les mêmes informations, cochez le mode *Activer l'optimisation des regroupements basés sur la variable cible effectués par K2C*.
> - Lorsque le mode *Utilise le codage naturel* est activé, seule la version d'origine de la variable est utilisée. Les versions codées sont désactivées et les critères d'exclusion sont relâchés pour les versions d'origine.

Travailler sans aucune structure définie

Si vous laissez la structure comme indéfinie, l'application détermine automatiquement le regroupement des catégories en fonction de leur interaction avec la variable cible à l'aide d'un codeur cohérent. Vous pouvez configurer deux paramètres dans ce cas :

- le nombre de segments pour les variables continues,
- l'optimisation des regroupements pour toutes les variables.

Nombre de segments pour les variables continues

Lorsque vous ne travaillez avec aucune structure définie, vous pouvez définir le nombre de segments pour les variables continues. Les valeurs permises pour ce paramètre sont comprises entre 1 et 20.

La population est ainsi découpée en autant de segments de taille similaire. Ces segments permettent de constituer des statistiques descriptives, en particulier la répartition des variables cible pour chaque segment, ce qui influe sur le codage de la variable par rapport aux variables cible.

Le nombre de segments a une influence sur le calcul de la capacité prédictive (KI) : plus les segments sont nombreux, plus le calcul du KI est précis pour la variable explicative. Néanmoins, cette influence est très faible.

- Pour définir le nombre de segments pour les variables continues
  1. Cliquez avec le bouton droit de la souris sur la ligne de la variable continue à modifier.
  2. Sélectionnez *Définir une structure*.
  3. Sélectionnez *Modifier le nombre de segments pour les variables continues*.
  4. Choisissez parmi les options présentées ci-dessous :

Si vous souhaitez modifier le nombre de segments pour...	alors...
toutes les variables continues du modèle	• Saisissez le nombre de segments voulus dans le champ en bas du panneau • Cliquez sur Affecter le même nombre de segments pour toutes les variables. • Cliquez sur OK.
Pour modifier le nombre de segments pour la variable en cours de modification.	• Saisissez le nombre de segments voulu dans la colonne Nombre de segments en haut du panneau. • Cliquez sur OK.

Optimisation des regroupements

Lorsque vous travaillez avec une structure définie et que vous souhaitez conserver vos catégories telles quelles pour la construction du modèle, vous devez désactiver cette option.

Dans le cas contraire ou si vous travaillez sans aucune structure définie, *Activer l'optimisation des regroupements pour toutes les variables* permet dans un grand nombre de cas, d'augmenter la robustesse du modèle (KR) avec une perte d'information minime (KI). Quand c'est possible, les segments adjacents similaires sont regroupés pour réduire les artéfacts observés entre les jeux de données d'entraînement et de validation.

- Pour activer l'optimisation des regroupements pour toutes les variables
  1. Effectuez un clic droit sur la ligne correspondant à la variable à modifier.
  2. Sélectionnez *Définir une structure*.
  3. Sélectionnez *Activer l'optimisation des regroupements pour toutes les variables*, de façon à ce que l'option soit cochée.

5.3.3.7 Ajout d'un filtre au jeu de données

Si vous ne souhaitez pas utiliser le jeu de données entier, vous pouvez créer un filtre pour ne retenir que les données souhaitées.

- Ajouter un filtre
  1. Cochez l'option *Ajouter un filtre au jeu de données*.

2. Cliquez sur le bouton *Suivant*.

5.3.3.8 Codage des variables

La création d'un modèle utilise non seulement les variables d'origine mais, en cas de variables continues ou ordinales, leur valeur est codée par Automated Analytics. Cette méthode s'intitule le *codage dual* et permet à Automated Analytics de trouver toutes les informations contenues dans chaque variable. Si vous activez le *codage naturel* pour une variable donnée, sa valeur codée (`c_variableName`) n'est pas générée.

- Pour activer le codage naturel des variables
  1. Dans le panneau de description des données, sélectionnez la variable que vous souhaitez coder naturellement, par exemple "age". Vous pouvez sélectionner plusieurs variables en appuyant sur la touche CTRL.
  2. Effectuez un clic droit et sélectionnez l'option *Codage* > *Utilise le codage naturel*.

5.3.4 Paramètres avancés de l'espace

- Échantillonnage [page 106]
- Paramètres des fichiers [page 107]
- Ligne d'en-tête [page 109]

5.3.4.1 Échantillonnage

Le jeu de données d'entraînement peut être échantillonné pour toutes les stratégies de partition, à l'exception de la stratégie *personnalisée*. Cliquez sur le bouton (*Paramètres avancés*) pour afficher l'écran *Échantillonnage*.

L'échantillonnage peut être réalisé de la façon suivante :

- Sélection d'une première ligne et d'une dernière ligne spécifiques dans les données originales.
- Sélection aléatoire d'une proportion des données originales. Une graine aléatoire permet de sélectionner les mêmes enregistrements dans les données originales lors de l'échantillonnage.
- Utilisation simultanée de la sélection de ligne et de l'échantillonnage aléatoire.
- Pour paramétrer un échantillonnage aléatoire
  1. Cochez l'option *Sélection aléatoire*.
  2. Dans la section *Sélection aléatoire*, utilisez le curseur *Proportion* pour sélectionner le pourcentage du jeu de données d'origine que vous souhaitez conserver.
  3. Dans le champ *Graine aléatoire*, saisissez une valeur pour l'échantillonnage. La valeur par défaut de la graine est 1960.

4. Dans la partie *Proportion*, utilisez le curseur pour sélectionner la proportion de l'observation qui doit apparaître dans le jeu de données. La sélection aléatoire génère un index aléatoire pour chaque ligne du jeu de données basé sur la graine. Une même graine pour un même index.
   Vous pouvez affiner la sélection en cochant l'option *Avancé*. Le curseur est remplacé par deux champs qui vous permettent d'indiquer un intervalle correspondant aux index à conserver. Par exemple, si vous sélectionnez [0, 20 %], vous sélectionnez les lignes avec un index compris entre 0 et 0,2. Ainsi [0, 20 %] ne présente pas les mêmes lignes que [80 %, 100 %](index compris entre 0,8 et 1). En réalité, si vous souhaitez apprendre sur 80 % de votre jeu de données, appliquer sur les 20 % restants et effectuer un échantillonnage aléatoire, c'est la seule méthode. Vous pouvez alors créer deux jeux de données complémentaires en conservant les mêmes paramètres de curseur et de graine et en sélectionnant deux intervalles complémentaires.

- Pour paramétrer une sélection par lignes
  1. Cochez l'option *Sélection par lignes*.
  2. Dans le champ *Première ligne*, saisissez le numéro de la ligne par laquelle commencera le jeu de données échantillonné.
  3. Dans le champ *Dernière ligne*, saisissez le numéro de la ligne par laquelle se terminera le jeu de données échantillonné.
  4. Une fois l'échantillonnage spécifié, cliquez sur *OK* pour enregistrer les paramètres et fermer la fenêtre *Paramètres*.

5.3.4.2 Paramètres des fichiers

Définition du séparateur

Lorsque la source de données est un fichier texte, ce panneau permet également de choisir le séparateur de colonnes. Par défaut, Automated Analytics détermine le séparateur en analysant les cents premières lignes de données. Le séparateur déterminé par l'application est sélectionné lorsque vous ouvrez la fenêtre des paramètres avancés.

- Pour définir le séparateur
  1. Cochez l'option *Définir le séparateur*.
  2. Sélectionnez le séparateur correspondant à votre fichier texte. Quatre séparateurs prédéfinis sont disponibles : *tabulation*, *virgule*, *point-virgule* et *espace*. Vous pouvez néanmoins définir votre propre séparateur en sélectionnant l'option *Autre* et en entrant la valeur de séparateur du fichier texte.
  3. Cliquez sur le bouton *OK* pour enregistrer les paramètres.

Format de date

Cette section vous permet de spécifier le format de date utilisé dans le fichier de données. Par défaut, Automated Analytics utilise le format ISO *AAAA-MM-JJ HH:MM:SS*.

- Pour spécifier le format de date
  1. Cochez l'option *Format de date*.
  2. Dans la liste *Ordre de la date*, choisissez le format de date.
  3. Dans le champ *Séparateur*, saisissez les caractères utilisés pour séparer les différents éléments de date.
     Une note affichée sous le champ indique si le séparateur est valide et montre un exemple.

Sélection du codage

Cette section vous permet de choisir le codage des caractères utilisé pour lire le fichier de données. Par défaut, Automated Analytics considère que le fichier est au format *natif*.

- Pour choisir le codage des caractères
  1. Cochez l'option *Choisir le codage des caractères*.
  2. Sélectionnez le codage approprié parmi les quatre options disponibles.

Sélection du format des angles

Cette option vous permet de définir le format des angles. Dans ce cas, les angles font référence à la latitude et à la longitude, qui sont les angles utilisés comme coordonnées géographiques lors de la définition d'un point sur une carte. Ces angles peuvent être indiqués dans différents formats.

Le tableau suivant récapitule les formants des angles proposés :

Nom du format des angles	Format	Exemple
Degrés décimaux	[-+]ddd.ddddd	+32,30642 -122,61458
Degrés minutes décimaux	[-+]ddd° mm.mmmm'[SNWE]	+32° 18.385' -122° 36.875'
Degrés 100 minutes décimaux (NMEA)	[-+]dddmm.mmm	+3218,3852 -12236,8748
Degrés minutes secondes	[-+]ddd° mm' ss.ss"	+32° 18' 23.1" -122° 36' 52.5"

- Pour spécifier le format des angles
  1. Cochez l'option *Spécifier le format des angles*.
  2. Sélectionnez dans la liste le format des angles souhaité.
  3. Cliquez *OK* pour valider.

5.3.4.3 Ligne d'en-tête

Cette section vous permet de spécifier la ligne de votre fichier de données à utiliser comme ligne d'en-tête. Par défaut, Automated Analytics analyse les cent premières lignes des fichiers et sélectionne comme ligne d'en-tête la première ligne contenant le nombre de tabulations le plus récurrent. Tout ce qui se trouve au-dessus de cette ligne est considéré comme des commentaires.

- Pour forcer la ligne d'en-tête
  1. Cochez l'option *Forcer la ligne d'en-tête*.
  2. Dans le champ *Ligne*, entrez le numéro de la ligne à utiliser en tant que ligne d'en-tête.

5.3.5 Stratégie de partition

Les fonctionnalités SAP Predictive Analytics utilisent deux sous-ensembles de données obligatoires et un facultatif pour la modélisation :

- *Entraînement* : utilisé pour l'entraînement des modèles.
- *Validation* : utilisé pour garantir la cohérence (robustesse) des modèles.
- *Test* (facultatif) : sur lequel seront calculés certains estimateurs de qualité du modèle.

Vous disposez de neuf stratégies de partition pour générer vos modèles :

- Une stratégie de partition personnalisée pour laquelle chaque sous-ensemble de données peut être indiqué explicitement.
- Huit stratégies de partition automatiques :
  - *Séquentielle* : les trois sous-ensembles seront générés séquentiellement à partir des données originales (3/5 pour l'entraînement, 1/5 pour la validation et 1/5 pour le test). Il s'agit de la stratégie de partition la plus rapide.
  - *Périodique* : les données d'origine sont distribuées périodiquement dans les sous-ensembles de données par groupes de cinq enregistrements : les trois premiers pour l'entraînement, le quatrième pour la validation et le cinquième pour le test. Cette stratégie est recommandée si l'ordre des données originales doit être ignoré.
  - *Aléatoire* : les données d'origine sont distribuées de manière aléatoire dans les trois sous-ensembles. Cette stratégie est recommandée si les données sont triées selon une variable telle que l'âge ou selon la variable cible.
  - *Périodique avec test à la fin* : stratégie de partition périodique terminée par un test.
  - *Aléatoire avec test à la fin* : stratégie de partition aléatoire terminée par un test.
  - *Séquentielle sans test* : stratégie de partition séquentielle utilisant uniquement les jeux de données d'entraînement et de validation.
  - *Périodique sans test* : stratégie de partition périodique utilisant uniquement les jeux de données d'entraînement et de validation.
  - *Aléatoire sans test* (*Par défaut*) : stratégie de partition aléatoire utilisant uniquement les jeux de données d'entraînement et de validation.
- Pour sélectionner une stratégie de partition automatique
  1. Cochez l'option *Prédéfinie*.

2. Dans la liste, sélectionnez la stratégie de partition que vous souhaitez utiliser.
3. Cliquez sur le bouton *Suivant*. Le panneau *Données à modéliser* est affiché.

- Pour sélectionner la stratégie de partition personnalisée
  1. Cochez l'option *Personnalisée*.
  2. Utilisez le bouton *Parcourir* correspondant au champ *Entraînement* pour sélectionner le jeu de données à utiliser comme jeu de données d'entraînement.
  3. Utilisez le bouton *Parcourir* correspondant au champ *Validation* pour sélectionner le jeu de données à utiliser comme jeu de données d'e validation.
  4. Si vous souhaitez utiliser un jeu de données de test, utilisez le bouton *Parcourir* correspondant au champ *Test* pour sélectionner le jeu de données à utiliser. Sinon laissez ce champ vide.
  5. Cliquez sur le bouton *Suivant*. Le panneau *Données à modéliser* est affiché.

5.3.6 Aperçu des données

- Aperçu des données [page 110]
- Statistiques descriptives [page 111]
- Graphe [page 113]
- Variables de position [page 113]

5.3.6.1 Aperçu des données

Par défaut, Automated Analytics affiche les cent premières lignes du jeu de données, mais vous pouvez modifier ce nombre.

- Pour afficher les données :
  - Cliquez sur le bouton . La fenêtre *Aperçu des données* s'ouvre affichant les 100 premières lignes de votre jeu de données.
- Pour changer le nombre de lignes affichées :
  1. Dans le champ *Première ligne*, saisissez le numéro de la première ligne que vous souhaitez afficher.
  2. Dans le champ *Dernière ligne*, saisissez le numéro de la dernière ligne que vous souhaitez afficher.
  3. Cliquez sur le bouton *Actualiser* . Les lignes sélectionnées s'affichent dans la fenêtre.
- Pour afficher une variable spécifique :

Lorsque votre jeu de données contient un grand nombre de variables, une option vous permet de naviguer dans votre jeu de données pour visualiser directement une variable spécifique.

1. Cliquez sur le bouton de recherche de variables ( ) dans le coin supérieur droit de la fenêtre. La fenêtre *Rechercher une variable par* s'affiche.
2. Sélectionnez le mode de recherche : par *index* (le numéro de la variable dans le jeu de données) ou par *nom*.
3. Selon le mode choisi, saisissez l'index de la variable ou sélectionnez son nom dans la liste déroulante proposée.

4. Cliquez sur le bouton *OK*. La fenêtre de recherche se ferme et l'aperçu du jeu de données se recadre sur la variable choisie.

5.3.6.2 Statistiques descriptives

Cette option vous permet de calculer les statistiques pour :

- l'ensemble du jeu de données (option *Calculer les statistiques sur l'ensemble du jeu de données*) ;
- les n premières lignes du jeu de données (option *Calculer les statistiques sur les n premières lignes*) ;
- un sous-ensemble du jeu de données (option *Avancé...*).

La durée de calcul des statistiques dépend du volume des données à analyser : plus le nombre de lignes est élevé plus l'analyse sera longue.

Lorsque vous cliquez sur l'option de votre choix, la génération des statistiques est lancée. Une barre de progression vous indique l'avancement de l'analyse.

À la fin du processus d'analyse trois onglets s'affichent :

- *Variables*, qui contient la liste des variables trouvées dans le jeu de données avec leur valeur, leur type de stockage, le nombre de valeurs manquantes pour chaque variable ainsi que leur rôle lors de la modélisation.
- *Fréquence des catégories*, qui présente pour chaque variable l'ensemble de ses catégories, leur fréquence et le nombre d'occurrences de chaque catégorie (*Effectif pondéré*).
- *Variables continues*, qui présente pour chaque variable continue son minimum, son maximum, sa moyenne et son écart-type.

Options d'affichage

Il existe quatre modes d'affichage différents en fonction de l'onglet que vous consultez :

Bouton	Mode	Ce mode vous permet de ...
	Table	Trier les statistiques par colonne. Afficher un rendu graphique sous forme d'une vue concernant la colonne ou l'ensemble du jeu de données analysé. Ce rendu graphique peut être activé/désactivé en cliquant sur le bouton .
	Rapport HTML	Afficher les statistiques sous forme HTML.

Bouton	Mode	Ce mode vous permet de ...
	Diagramme à barres	Afficher les statistiques sous forme de diagramme à barres dont les colonnes peuvent être : • non triées ; • triées par ordre croissant ; • triées par ordre décroissant ; • triées par ordre alphabétique ; • triées par ordre alphabétique inversé • Sélectionner les séries à afficher • Choisir les couleurs d'affichage du diagramme à barres
	Diagramme à secteurs	Afficher les statistiques sous forme de secteur
	Changer l'orientation des barres	Changer l'orientation des barres d'horizontal à vertical et inversement.

Options de copie et d'enregistrement

Plusieurs options de copie et d'enregistrement sont proposées :

Bouton	Option	Cette option vous permet de...
	Copier la vue actuelle	Copier au format HTML les statistiques affichées à l'écran
	Copier toutes les vues	Copier en format HTML les statistiques affichées à l'écran
	Copier l'image	Copier au format image le graphe affiché à l'écran
	Enregistrer la vue actuelle	Enregistrer en format texte séparé par des tabulations les statistiques affichées à l'écran
	Enregistrer toutes les vues	Enregistrer en format texte séparé par des tabulations les statistiques affichées à l'écran

5.3.6.3 Graphe

1. Dans le champ *Première ligne*, saisissez le numéro de la première ligne que vous souhaitez afficher.
2. Dans le champ *Dernière ligne*, saisissez le numéro de la dernière ligne que vous souhaitez afficher.
3. Dans la liste *Variable de l'axe des abscisses*, sélectionnez la variable que vous voulez afficher sur l'axe des abscisses.
4. Dans la liste *Variables disponibles*, sélectionnez les variables que vous voulez afficher sur l'axe des ordonnées et utilisez la flèche pour les inclure dans la liste des *variables de l'axe des ordonnées*.
5. Cliquez sur le bouton *Afficher le graphe*.
   Par défaut, le graphe affiché est un graphe à point. Pour afficher le graphe sous forme de lignes, cliquez sur *Tracer la courbe*.
6. Cliquez sur *Paramètres du graphe* pour modifier les paramètres.

5.3.6.4 Variables de position

L'onglet *Variables de position* permet d'afficher dans Google Earth la variable composite que vous avez définie plus tôt.

1. Sélectionnez la variable composite que vous voulez afficher.
2. Cliquez sur *Prévisualiser dans Google Earth*.

5.4 Définition des paramètres

5.4.1 Sélection de variables

Pour être incluse dans un processus de modelage, une variable doit être affectée à un des rôles décrit ci-dessous. Les variables sont explicatives par défaut sauf si vous les affectez à une cible ou un rôle de poids ou que vous choisissez de les exclure.

Rôles des variables dans le processus de modelage

Rôle de la variable	Définition
Variable explicative	Une variable explicative est utilisée pour calculer le modèle.
Variables cible	La variable cible est celle que vous cherchez à expliquer ou dont vous voulez prédire les valeurs dans un jeu de données d'application.

Rôle de la variable	Définition
Variable de poids	Une variable de poids attribue un poids relatif à chaque ligne qu'il décrit et permet d'orienter activement la phase d'entraînement du processus de modélisation. Lorsque vous déclarez une variable de poids, vous créez une nouvelle ligne proportionnelle à chaque valeur décrite par la valeur initiale. Par exemple, dans un jeu de données contenant les commandes d'achat avec indication de la quantité d'articles achetés, si un client achète une blouse et trois jupes, la ligne contenant la blouse est traitée comme occurrence simple, mais la ligne concernant l'achat des jupes est dupliquée pour créer trois lignes au lieu d'une. Cela donne à la commande d'achat des jupes un poids reflétant son importance relative dans le jeu de données de manière plus réaliste que son occurrence simple précédente. Vous remarquerez que lorsque la valeur pour la variable de poids est manquante, l'application ignore la ligne. L'application ne permet pas d'utiliser une variable pour laquelle les valeurs multiples sont manquantes.
Variables exclues	Vous pouvez choisir d'exclure certaines variables de la phase de modélisation. L'exclusion des colonnes qui n'ont pas d'influence sur les variables cible, comme un numéro de compte, peut accélérer l'exécution tout en les conservant. Cela n'interfère pas avec le processus de modélisation. Cependant, vous devez exclure les variables qui sont directement liées aux variables cible, telles que les transformations de variables cible et les variables contenant indirectement les mêmes informations que la variable cible. Par exemple, si un jeu de données contient la variable "a acheté le produit Oui/Non", vous devez exclure la variable "Montant de la facture" si elle contient le prix du produit.

-
-
-

5.4.1.1 Variables cible

- Pour choisir les variables cible
  1. Sélectionnez les variables dans la liste *Variables explicatives sélectionnées*.
  2. Cliquez sur le bouton > à gauche de la section *Variables cible*.

> **i** Remarque
>
> Seules les variables nominales avec deux catégories ou continues peuvent être utilisées comme variables cible dans un modèle.

5.4.1.2 Variable de poids

La sélection d'une *variable de poids* permet de définir l'option *Poids de référence* dans les *paramètres avancés du modèle*.

- Pour choisir une variable de poids
  1. Sélectionnez la variable dans la liste *Variables explicatives sélectionnées*.
  2. Cliquez sur le bouton > à gauche de la section *Variable de poids*.

> **i** Remarque
>
> Seules des variables numériques peuvent être choisies comme variables de poids.

5.4.1.3 Variables exclues

- Pour exclure une variable
  1. Sélectionnez la variable à exclure dans la liste *Variables explicatives conservées*.
  2. Cliquez sur le bouton **>** à gauche de la section *Variables exclues* de l'écran.

> **i** Remarque
>
> Toutes les listes peuvent être triées par ordre d'apparition dans le fichier ou par ordre alphabétique.

Si vous utilisez le même jeu de données pour créer différents modèles, l'enregistrement de la liste de variables que vous ne souhaitez pas utiliser facilitera cette étape.

- Pour enregistrer une liste de variables exclues
  1. Cliquez sur en dessous de la section *Variables exclues* de l'écran. Recherchez un emplacement.
  2. Saisissez un nom pour le fichier enregistrez-le.
- Pour charger une liste de variables exclues
  1. Cliquez sur en dessous de la section *Variables exclues* de l'écran.
  2. Recherchez et sélectionnez le fichier enregistré précédemment.

5.4.2 Récapitulatif des paramètres de modélisation

Cet écran résume les paramètres de modélisation. Vous pouvez spécifier les paramètres suivants :

- le champ *Nom du modèle* est rempli automatiquement. Il correspond au nom de la variable cible, suivi d'un trait de soulignement ("_") et du nom de la source de données, sans son extension.
- la *Description* qui, une fois le modèle enregistré, permet de le reconnaître quand vous voulez le rouvrir avec SAP Predictive Analytics.

5.4.2.1 Pour tous les composants

Nom du modèle	Le nom du modèle est renseigné automatiquement. Il correspond au nom de la variable cible, suivi d'un trait de soulignement ("_") et du nom de la source de données, sans son extension.
Description	Une fois le modèle enregistré, la description vous permettra de reconnaître votre modèle lorsque vous voudrez le rouvrir à l'aide de SAP Predictive Analytics.
Modélisation	Ce champ non-modifiable indique quel type de moteur sera utilisé pour créer le modèle.
Sauvegarde automatique...	Ce bouton ouvre le panneau *Sauvegarde automatique* qui vous permet d'indiquer que le modèle doit être enregistré automatiquement à la fin du processus d'apprentissage et de définir tous les paramètres requis. Lorsque la sauvegarde automatique est activée, une coche verte s'affiche sur le bouton *Activer la sauvegarde automatique*. À la fin du processus d'apprentissage, un message d'avertissement s'affiche, confirmant que le modèle a bien été enregistré.
Exporter le script KxShell...	Ce bouton ouvre le panneau vous permettant de générer un script KxShell qui reproduit le modèle en cours.

5.4.2.2 Classification/Régression

- Calculer l'arbre de décision

Il est possible d'afficher les résultats générés sous la forme d'un arbre de décision. Pour activer cette option, cochez la case *Calculer l'arbre de décision*.

- Activer la sélection automatique des variables

La sélection automatique des variables est activée par défaut. Cette fonctionnalité vous permet de réduire automatiquement le nombre de variables du modèle par rapport à des critères de qualité. Vous pouvez modifier les critères de sélection dans le panneau *Paramètres spécifiques du modèle*. Pour la désactiver, décochez la case *Activer la sélection automatique des variables*.

- Avancé...

Le bouton *Avancé...* vous permet d'accéder à l'écran *Paramètres spécifiques du modèle*, où vous pouvez définir :

- les paramètres généraux (ordre du polynôme du modèle à générer, nombre de segments de score, poids de référence, paramètres des corrélations, paramètres de post-traitement et valeur clé de chaque variable cible),
- la sélection automatique des variables,
- les modes d'apprentissage spécifiques (mode Règle et mode Risque),
- la table de profit.

5.4.2.3 Mise en cluster

- Choisir le meilleur nombre de clusters dans cet intervalle

Ce champ vous permet de spécifier combien de clusters devront être générés par le modèle. Par défaut, le nombre de clusters est défini à 10. Plus le nombre de segments est élevé, moins le modèle est robuste (reproductibilité). Plus le nombre de segments est faible, moins les informations sont précises (capacité prédictive). Il est généralement conseillé de commencer par le nombre de clusters par défaut, puis de l'augmenter ou le réduire selon les résultats obtenus.

- Pour une *segmentation* (c'est-à-dire avec variable cible), l'utilisateur choisit le meilleur nombre de segments, par exemple [5;10], ce qui signifie que l'utilisateur souhaite avoir entre 5 et 10 clusters. Le moteur calcule le "meilleur nombre de clusters" à l'aide de la métrique Reproductibilité + Capacité prédictive. Vous pouvez, par exemple, avoir 7 clusters.
- Pour un *regroupement* (c'est-à-dire sans variable cible), l'application choisit le nombre minimum de clusters, par exemple [10;10], ce qui signifie que l'utilisateur souhaite avoir 10 clusters.

> **i** Remarque
>
> Lorsque vous activez l'option *Calculer les expressions SQL*, un cluster supplémentaire contenant les enregistrements non assignés est généré.

- Calculer les expressions SQL

Cette fonctionnalité vous permet de voir les expressions SQL utilisées pour générer chaque cluster. Le calcul des expressions SQL est activé par défaut.

- Avancé...

Le bouton *Avancé* vous permet d'accéder au panneau. Ce panneau vous permet d'activer les statistiques croisées, de spécifier la valeur clé de chaque variable cible ainsi que la distance à utiliser et la stratégie de codage.

5.4.2.4 Série temporelle

- Nombre de prévisions
  L'unité de temps indiquée dans le champ *Nombre de prévisions* dépend des données analysées. Par exemple, si les observations du jeu de données sont enregistrées mensuellement, l'unité de temps sera le mois.
- Prévision maximum
  Ce champ indique l'horizon maximal de prévision pour lequel il y a suffisamment de variables extra-prédictibles dans le jeu de données.

5.4.3 Définition de paramètres pour la classification/ régression

Cette section décrit les paramètres avancés spécifiques aux modèles de classification ou de régression.

5.4.3.1 Paramètres avancés du modèle

Lorsque tous les paramètres avancés sont définis, cliquez sur le bouton *Valider* pour enregistrer vos paramètres et revenir au panneau *Récapitulatif des paramètres de modélisation*.

- Général [page 118]
- Sélection automatique des variables [page 122]
- Mode d'apprentissage [page 124]
- Table de profit [page 125]

5.4.3.1.1 Général

Ordre du polynôme

Les modèles construits par Automated Analytics Modeler - Régression/Classification sont des expressions polynomiales des données d'entrée. Les modèles produisent une estimation de variable cible comme fonction des variables d'entrée X1,X2,...Xn. L'ordre est le degré maximum autorisé dans une telle expression.

Le seul paramètre disponible pour le modèle est l'ordre du polynôme. On commence généralement par un modèle d'ordre 1, puis on avance vers des modèles d'ordre plus élevé. L'ordre du polynôme est limité par :

- le nombre de variables d'entrée ;
- la capacité mémoire de l'ordinateur ;
- le temps accordé à la phase de modélisation.

Description du polynôme

Un polynôme peut être de degré 1, 2, 3 ou plus. En définissant l'ordre du polynôme, vous définissez le degré de complexité du modèle.

Exemples de polynômes

- Un polynôme de degré 1 a la forme :
  ```
  Y = a1.X1 + ... + an.Xn
  ```
- Un polynôme de degré 2 a la forme :
  ```
  Y = A1.X1 + ... + An.Xn +
  B1.X1.X1 + B2.X2.X2+ ...Bk.Xn.Xn +
  ```

```
C1.X1.X2 + ... + Cm.X[n-1].Xn
```

Méthodologie

Dans la majorité des cas, un polynôme de degré 1 est suffisant pour générer un modèle pertinent et robuste. Utiliser un polynôme de degré plus élevé ne garantit pas forcément de meilleurs résultats que ceux obtenus par un polynôme de degré 1. De plus, plus le degré du polynôme est élevé :

- plus le modèle correspondant est long à générer ;
- plus l'application du modèle à de nouveaux jeux de données va prendre de temps ;
- plus il est difficile d'interpréter les résultats de la modélisation.

La sélection de l'un ou l'autre des degrés du polynôme dépend de la nature des données à analyser. La méthode recommandée est la suivante :

1. Commencez par créer un modèle avec un degré d'ordre. Dans la grande majorité des cas, ce degré d'ordre sera suffisant pour garantir un modèle pertinent et robuste.
2. Testez ensuite les résultats ainsi obtenus avec un modèle de degré plus élevé si les performances du modèle de degré 1 semblent insuffisantes.

Nombre de classes pour la variable de score

Cette option vous permet de définir le nombre de classes à créer pour la variable score. La valeur saisie doit être entre 20 et 100 ; en effet un nombre de classe inférieur ou supérieur nuirait à la qualité du modèle.

Exclusion des variables à faible KR

Cette option vous permet d'activer l'exclusion de variables en fonction de la valeur de leur reproductibilité (KR). Pour déterminer si une variable présente une faible reproductibilité, Automated Analytics utilise un seuil calculé en interne qui dépend principalement de la taille du jeu de données et de la distribution de la cible.

L'exclusion automatique des variables à faible reproductibilité est désactivée par défaut. Si vous n'activez pas cette fonctionnalité, aucune variable ne sera exclue à cause de la valeur de sa reproductibilité.

- Pour exclure automatiquement les variables à faible reproductibilité :
  - Cochez l'option **Exclusion des variables à faible reproductibilité**.

Poids de référence

La sélection d'une *variable de poids* vous permet d'utiliser l'option *Poids de référence*.

La notion de *Poids de référence* a été ajoutée afin de définir un seuil en dessous duquel une catégorie sera associée à KxOther.

Les *rapports statistiques* incluent maintenant les informations sur les poids dans les *statistiques descriptives* et la *taille des jeux de données*.

- Pour définir un Poids de référence :
  1. Cochez la case *Poids de référence*.
  2. Entrez une valeur pour le seuil. La valeur par défaut est 1.

Paramètres des corrélations

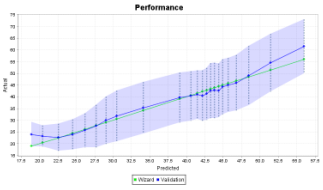
Cette option vous permet de paramétrer le nombre de corrélations à afficher dans le panneau de débriefing *Corrélations*. Autrement dit de choisir combien de corrélations devront être affichées dans ce panneau.

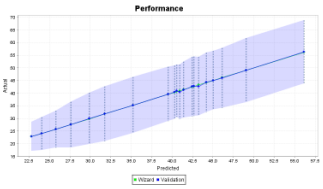
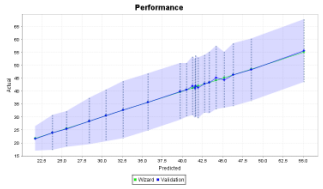
Deux variables hautement corrélées contiennent les mêmes informations par rapport à la variable cible. À chaque corrélation correspondent donc deux variables et un taux de corrélation. Lorsque vous modifiez le nombre de corrélations à afficher, le moteur supprime celles dont le taux de corrélation est le moins élevé, conservant ainsi uniquement les plus significatives.

Post-traitement

Cette section vous permet de paramétrer la régression selon trois stratégies. Cette option ne peut être activée que si le modèle contient au moins une variable cible continue.

Le tableau ci-dessous fournit la description de ces stratégies et un exemple de courbe de performance pour chaque stratégie.

Stratégie de régression	Description	Exemple de courbe de performances
Pas de post-traitement	La première stratégie consiste à désactiver le post-traitement de la régression au cours de la phase du modèle d'apprentissage afin de créer une régression similaire à celle utilisée dans les versions antérieures à 3.3.2. Dans ce cas, une régression standard est exécutée. Aucune optimisation spécifique n'est appliquée aux scores finaux. Les valeurs cible d'origine sont utilisées et les valeurs de score brutes sont générées en sortie.	

Stratégie de régression	Description	Exemple de courbe de performances
Codage original de la cible	La seconde consiste à activer la phase de redressement des prédictions et à utiliser la valeur de la cible directement lors de l'apprentissage du modèle pour calculer les coefficients de régression. Pendant la phase de post-traitement, le résultat de la régression est ensuite transformé afin d'aligner les moyennes des segments du score à celles de la variable cible. Remarque : C'est la stratégie utilisée par défaut dans Automated Analytics.	
Codage uniforme de la cible	La dernière stratégie qui s'applique aux régressions utilisant un post-traitement consiste à utiliser d'abord une valeur cible codée à la place de la valeur cible d'origine lors de la phase d'apprentissage du modèle afin d'obtenir une distribution uniforme : c'est la phase de prétraitement. Ensuite, les coefficients de régression sont calculés et les scores sont convertis dans l'espace d'origine de la cible au cours de la phase de post-traitement. Remarque : Cette stratégie est à privilégier lorsque la stratégie par défaut ne produit pas des modèles de qualité satisfaisante, ce qui est souvent le cas avec des distributions dissymétriques des valeurs de cible.	

- Régression sans post-traitement
  - Décochez la case *Activer le post-traitement*.

    > **i** Remarque
    >
    > Il est impossible de modifier la stratégie de codage cible lorsque le post-traitement est désactivé.

- Régression utilisant des valeurs cible d'origine
  1. Cochez la case *Activer le post-traitement*.
  2. Sélectionnez la case d'option *Codage d'origine de la cible*.
- Régression utilisant le codage uniforme de la cible
  1. Cochez la case *Activer le post-traitement*.
  2. Sélectionnez la case d'option *Codage uniforme de la cible*.

Paramètres des clés cible

La section *Définir la valeur des clés cible* énumère les variables cible sélectionnées dans l'écran *Sélection des variables* et permet de choisir leur valeur clé. Pour sélectionner la valeur clé des variables cible, saisissez la valeur dans le champ *Clé cible* correspondant à chaque cible. Par défaut, la valeur clé paramétrée automatiquement est la catégorie la moins fréquente de la variable cible.

5.4.3.1.2 Sélection automatique des variables

Cette section vous permet de réduire automatiquement le nombre de variables du modèle par rapport à des critères de qualité. Cette sélection se fait par itérations successives. Il existe deux modes de sélection, l'un repose sur le nombre de variables à conserver, et l'autre sur la quantité d'information à conserver. La quantité d'information correspond à la somme des contributions des variables.

- Pour utiliser la sélection automatique des variables :
  - Cochez la case *Activer la sélection automatique des variables*. Les options correspondantes sont activées.
    Par défaut, les paramètres sont définis sur : *"Sélectionner le meilleur modèle en conservant entre 1 et toutes variables."*
  - Chaque paramètre modifiable est signalé sous forme de lien hypertexte (bleu, souligné).

Mode de sélection

- Pour choisir le mode de sélection :
  1. Cliquez sur le lien caractérisant le type d'information à conserver à chaque itération du processus de sélection. Par exemple, *le meilleur modèle* dans la phrase "*Sélectionner le meilleur modèle en conservant entre 1 et toutes variables*".
     Une liste déroulante s'affiche proposant les options suivantes :
     - *le meilleur modèle*
     - *le dernier modèle généré*
  2. Sélectionnez l'option de votre choix.
  3. Cliquez sur le bouton *OK*.
- Pour sélectionner le nombre de variables :

Ce critère est obligatoire et permet de fixer le nombre minimum et le nombre maximum de variables du modèle final.

1. Dans la phrase définissant le nombre de variables "*Sélectionner le meilleur modèle en conservant entre 1 et toutes les variables*", sélectionnez le nombre minimum de variables (par exemple, *1 variable*) et le nombre maximum de variables (par exemple, *toutes les variables*).
2. Pour le nombre minimum de variables, une règle allant de 1 au nombre total de variables dans le modèle s'affiche.
   Déplacez le curseur pour sélectionner la quantité de votre choix.
   Pour le nombre maximum de variables, vous pouvez soit confirmer le nombre minimum sélectionné auparavant en choisissant "*Conserver toutes les variables*", soit choisir un nombre maximum de variables.

3. Cliquez sur le bouton *OK*.

Paramètres de sélection des variables

Vous avez le choix entre deux paramètres de sélection des variables :

- *Chaque étape retire 1 variable*.

Cette option vous permet de paramétrer le nombre de variables qui devraient être exclues à chaque itération.

- *Chaque étape conserve 95,0 % de l'information*.

Cette option vous permet de paramétrer la quantité d'information qui devrait être conservé à chaque itération, limitant ainsi la perte d'information.

Sélectionnez l'option de votre choix.

- Pour sélectionner le nombre de variables :
  1. Cliquez sur le lien indiquant le nombre de variables dans la phrase "*Chaque étape supprime 1 variable.*" Une règle allant de 1 au nombre total de variables dans le modèle s'affiche.
  2. Déplacez le curseur sur la règle pour sélectionner le nombre de votre choix.
  3. Cliquez sur le bouton *OK*.
- Pour paramétrer la quantité d'information :
  1. Cliquez sur le lien indiquant la quantité d'information à conserver dans la phrase "*Chaque étape conserve 95.0 % d'information.*" Une règle s'affiche.
  2. Déplacez le curseur pour sélectionner la quantité de votre choix.
  3. Cliquez sur le bouton *OK*.
- Pour paramétrer la perte de qualité autorisée :

La perte de qualité est paramétrée dans la phrase "*La recherche s'arrête en cas de diminution de 1,0 % du KI et du KR*".

1. Cliquez sur le lien indiquant le pourcentage de perte (par exemple, *5,0 %*). Une règle s'affiche.
2. Sélectionnez le pourcentage maximal autorisé de perte de qualité.
3. Cliquez sur *OK*.
4. Cliquez sur le critère de qualité. Une liste déroulante s'affiche proposant les options suivantes :
   - *Basé sur KI + 2KR,* la perte de qualité est basée sur les deux indicateurs KI (apport d'information du modèle) et 2 KR (robustesse).
   - *KI et KR*, la perte de qualité est limitée pour les deux indicateurs KI (apport d'information du modèle) et KR (robustesse). C'est la valeur par défaut.
   - *KI*, la perte de qualité est seulement limitée pour le KI.
   - *KR*, la perte de qualité est seulement limitée pour le KR.
5. Sélectionnez l'option de votre choix.
6. Cliquez sur le bouton *OK*.

5.4.3.1.3 Mode d'apprentissage

Cet onglet vous permet de sélectionner un mode d'apprentissage spécifique pour votre modèle.

- Pour sélectionner un mode d'apprentissage spécifique :
  1. Cochez la case *Activer un mode d'apprentissage spécifique*. L'onglet est activé et les paramètres du *mode Risque* sont affichés.

Activer le mode Risque

Le *mode Risque* permet aux utilisateurs avancés de demander à un modèle de classification de traduire les équations internes qu'il a obtenues sans contrainte en une échelle de scores spécifiée associées au rapport bons/mauvais.

Quand ce mode est activé, les différents codages internes des variables continues et ordinales sont rassemblés en une seule représentation qui permet une vision simplifiée des équations internes du modèle. C'est particulièrement intéressant lorsque l'utilisation de modèles prédictifs est soumise à des restrictions légales : les équations du modèle sont désormais assez simples pour être comprises par les services juridiques et peuvent être présentées, non seulement dans un langage de programmation comme avant, mais également en termes simples.

La technologie sous-jacente est également utilisée pour afficher les "scorecard".

Pour utiliser ce mode, vous devez choisir :

- Un *score de risque* associé à un ratio des **probabilités bon/mauvais**

  > **i** Remarque
  >
  > Le ratio de probabilités est le rapport entre "bon" et "mauvais", c'est-à-dire (1-p)/p, p représentant la probabilité de risque

- Le *nombre de points pour doubler le rapport*

  > **i** Remarque
  >
  > Il s'agit du nombre de points de risque nécessaire pour doubler le ratio de probabilités.

  > **i** Remarque
  >
  > Prenons, par exemple un *score de risque* égal à 615, un ratio de probabilités de 9 :1 et 15 points pour doubler le rapport. Dans ce cas, Automated Analytics va ré-échelonner automatiquement les scores internes en scores dans l'espace *Mode risque* et associe un ratio de probabilités à chaque score dans l'espace *Mode risque*.

- Pour définir les paramètres du mode Risque :
  1. Dans le champ *Score de risque*, saisissez le score que vous voulez associer à un rapport bon/mauvais.
  2. Dans le champ *pour un rapport bon/mauvais de*, saisissez le rapport.
  3. Dans le champ *Points pour doubler le rapport*, indiquez le nombre de points dont le score doit augmenter pour doubler le rapport.

4. Cliquez sur le bouton *Voir la table de score* pour afficher un tableau des scores associés aux rapports bon/mauvais correspondants.

Domaine d'ajustement des risques

Cette option permet à l'utilisateur de paramétrer la manière dont l'ajustement des scores de risque est effectué, c'est-à-dire comment Automated Analytics ajuste ses propres scores aux scores de risque.

L'option d'ajustement des scores a deux modes :

- *Basé sur les points pour doubler le rapport* : l'aire d'ajustement des scores est égale à [Score médian - N*PDR ; Score médian + N*PDR]. N (nombre de points pour doubler le rapport autour du score médian) doit être spécifié par l'utilisateur. Par défaut, N est fixé à 2.

  > **i** Remarque
  >
  > PDR est l'abréviation de "Points pour doubler le rapport".

- *Basé sur la fréquence* : l'aire d'ajustement des scores est égale à [Quantile(Freq) ; Quantile(1.0 - Freq)]. La fréquence des scores extrêmes à exclure doit être spécifiée par l'utilisateur. Par défaut, elle est égale à 15 %.

Si vous ne cochez pas la case *Domaine d'ajustement des risques*, le mode *Basé sur la fréquence* sera utilisé par défaut.

L'ajustement des scores peut être pondéré.

- Pour paramétrer l'ajustement des risques :
  1. Cochez la case *Domaine d'ajustement des risques*.
  2. Sélectionnez le mode que vous voulez utiliser.
  3. En fonction du mode sélectionné, définissez la valeur appropriée dans le champ correspondant.
  4. Si vous voulez utiliser des poids pour l'ajustement, cochez la case *Utiliser les segments de la variable de score comme des poids*.

5.4.3.1.4 Table de profit

Cet onglet vous permet de calculer la table de profit sur le jeu de données d'entraînement, c'est-à-dire de classer vos données dans l'ordre des scores descendants, et de le fractionner en quantiles exacts (décile, vingtile, percentile). Cette option peut être utile pour étudier les performances du modèle à l'aide du jeu de données de validation.

- Pour calculer la table de profit :
  1. Cochez la case *Calculer la table de profit sur les données d'entraînement*.
  2. Dans la liste, sélectionnez le *nombre de quantiles* dans lequel vous voulez segmenter vos données.
  3. Vous pouvez ajouter des variables supplémentaires pour estimer le profit pour chaque segment de la population :
     1. Dans la liste *Variables*, sélectionnez les variables que vous souhaitez ajouter à la table de profit. Utilisez la touche CTRL pour sélectionner plusieurs variables.

2. Cliquez sur le bouton > pour ajouter les variables sélectionnées à la liste *Valeurs pour table de profit*.
   La somme de chaque variable sélectionnée sera calculée pour chaque segment de la population.

4. Cliquez sur le bouton *Valider* pour enregistrer les paramètres avancés et revenir au panneau *Récapitulatif des paramètres de modélisation*.

> **i** Remarque
>
> Le résultat du calcul de la table de profit est disponible après génération du modèle dans la section *Performance du modèle* des *rapports statistiques*.

5.4.4 Définition de paramètres pour la segmentation/mise en cluster

Cette section décrit les paramètres avancés spécifiques aux modèles de segmentation ou de mise en cluster.

5.4.4.1 Paramètres spécifiques du modèle

- Statistiques croisées [page 126]
- Clés cible [page 128]
- Stratégie de codage [page 129]

5.4.4.1.1 Statistiques croisées

L'option *Calculer les statistiques croisées* vous permet de visualiser le profil de chaque variable explicative pour chaque cluster, par rapport à leur profil pour le jeu de données entier.

Différence entre statistiques croisées classiques et expressions SQL

La segmentation créée avec les expressions SQL est différente de celle créée sans. La raison d'être des expressions SQL est de proposer des segments faciles à comprendre et à appliquer. Les expressions SQL doivent se rapprocher au plus près des segments de base (c'est-à-dire ceux que vous obtenez sans calculer les expressions SQL). Les expressions SQL peuvent être utilisées pour avoir une meilleure définition/compréhension des clusters et les déployer dans la base de données complète ou sur de nouvelles données (ce qui n'est généralement pas anodin avec les autres techniques).

L'utilisation de graphes permet de mieux comprendre la différence entre les clusters créés à partir de centroïdes et ceux créés à partir d'expressions SQL.

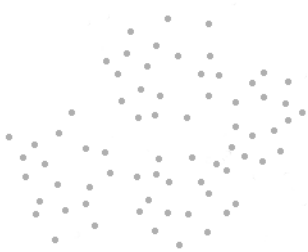
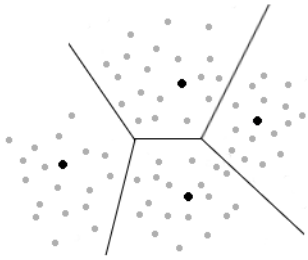
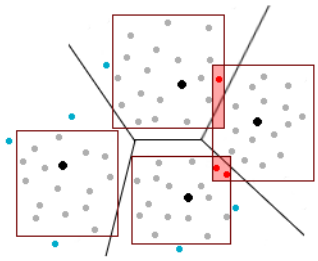
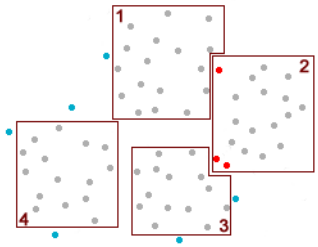
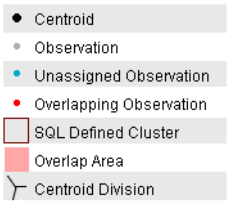
Schéma	Explication
	Ce schéma représente un ensemble d'observations issues d'un jeu de données.
	Pour créer un cluster, Automated Analytics Modeler - Segmentation/Mise en cluster utilise l'approche centroïde. Les centroïdes sont le résultat d'un algorithme de mise en cluster. Ils représentent ainsi le barycentre des points dont ils sont le plus proche. Quand on applique la segmentation à ce jeu de données, les observations sont regroupées en fonction de leur distance par rapport à chaque centroïde. Ce graphe représente le jeu de données regroupé en quatre clusters. Ce schéma est connu sous le nom de diagramme de Voronoï.
	Pour créer les expressions SQL qui définissent les clusters, Clustering utilise ce qu'on appelle la longueur minimum de description (Minimum Description Length ou MDL). Cela signifie que les clusters initiaux créés à partir de l'approche centroïde sont retravaillés pour correspondre à l'expression la plus simple possible essayant ainsi de trouver le meilleur compromis entre la taille de l'expression et la perte d'information. Ce graphe représente les expressions SQL des clusters (en rouge) comparées aux centroïdes. Sur ce graphe vous pouvez voir que : • certaines observations qui se trouvaient dans un cluster avec l'approche centroïde se retrouvent dans un autre quand on utilise les expressions SQL. • d'autres observations ne peuvent être décrites par les expressions SQL et sont donc laissées hors des clusters. Elles sont appelées observations non assignées. • certaines observations peuvent être décrites par deux expressions SQL distinctes et donc apparaître dans deux clusters différents. Cela s'appelle le *chevauchement*.

Schéma	Explication
	Ce graphe représente le résultat final obtenu avec les expressions SQL. Une observation ne peut pas apparaître dans deux clusters différents, donc lorsque deux clusters se recoupent, l'observation concernée est attribuée au premier cluster créé. Le second cluster qui contenait également l'observation est redéfini de manière à l'exclure. Dans ce schéma, les nombres correspondent à l'ordre de création des clusters. Vous constatez que les observations, qui se trouvaient dans deux clusters, sont conservées dans un seul cluster. Le choix du cluster dans lequel les observations seront conservées dépend de l'ordre dans lequel les règles SQL sont appliquées. Dans ce cas, la règle définissant le cluster 2 a été appliquée avant celles définissant les clusters 1 et 3.
	Légende des schémas

Comment choisir le type de segmentation le plus adapté ?

Grâce à la segmentation supervisée, Automated Analytics Modeler - Segmentation/Mise en cluster vous propose des indicateurs de performance : la capacité prédictive (KI) et la reproductibilité (KR). Le KI vous permet de comparer les deux segmentations (principalement parce que le nombre de segments est identique). Si le KI ne change pas de manière significative, la segmentation SQL est à privilégier car elle est plus simple à appréhender. Si le KI baisse, il vaut mieux en rester à la segmentation de base.

La capacité prédictive n'est pas forcément ce que vous voulez optimiser pour une segmentation. Vous pouvez visualiser le profil cible de chaque segment dans l'interface graphique. Parmi les quatre clusters, un ou deux peuvent être particulièrement intéressants. Dans ce cas, il vaut mieux se concentrer sur ces segments et étudier leur évolution lors de la génération des expressions SQL.

5.4.4.1.2 Clés cible

La section *Définir la valeur des clés cible* énumère les variables cible sélectionnées dans l'écran *Sélection des variables* et permet de choisir leur valeur clé.

- Pour sélectionner la valeur clé des variables cible

- Dans le champ *Clé cible* correspondant à chaque variable cible, saisissez la valeur choisie. Par défaut, la valeur de la catégorie cible est automatiquement fixée à la catégorie la moins fréquente de la variable cible.

5.4.4.1.3 Stratégie de codage

L'option *Stratégie de codage* fait référence au type de codage que le moteur de segmentation attend de l'encodeur de données d'Automated Analytics.

- Pour sélectionner une stratégie de codage :
  - Choisissez parmi les options suivantes dans la liste déroulante :

Option	Description
Automatique	Le système sélectionne automatiquement le meilleur codage en fonction des paramètres du modèle. Le codage *Moyenne de la cible* est utilisé pour les modèles de segmentation. Sinon, les variables sont codées à l'aide du schéma *Regroupement*.
Moyenne de la cible	**Valeur par défaut pour la segmentation** Chaque valeur de variable d'entrée continue est remplacée par la moyenne de la cible pour le segment auquel la valeur appartient. Chaque catégorie de variable d'entrée nominale est remplacée par la moyenne de la cible pour cette catégorie. Dans le cas d'une variable cible nominale, la moyenne de la cible correspond au pourcentage de cas positifs de la variable cible pour la catégorie de variable d'entrée.
Uniforme	Chaque segment de variable est codé dans l'intervalle `[-1;+1]` pour une distribution uniforme des variables.
Regroupement	**Valeur par défaut pour le regroupement** Une stratégie sans cible. Seule la fréquence du segment est utilisée pour coder les variables.

Les options suivantes ne sont disponibles que lorsque toutes les variables sont continues :

Option	Description
Natural	Aucune transformation n'est appliquée aux données d'entrée.
Min Max	Les variables sont codées dans l'intervalle [0,1], où 0 correspond à la valeur minimale de la variable et 1 à sa valeur maximale.

Option	Description
Normalisation de l'écart-type	Cette option applique une normalisation reposant sur la moyenne de la variable et l'écart-type. $\frac{x - Mean}{StdDev}$

5.4.5 Définition de paramètres pour les séries temporelles

Cette section décrit les paramètres avancés spécifiques aux modèles de séries temporelles.

5.4.5.1 Sélection de variables

Cet écran vous permet de :

- sélectionner les *variables requises* (date et cible) ;
- sélectionner une variable de *poids* ;
- exclure des variables de la modélisation ;
- réduire le jeu de données.

- Variables requises [page 130]
- Variable de poids [page 131]
- Affichage du signal [page 131]
- Variables exclues [page 132]
- Réduction du jeu de données [page 132]

5.4.5.1.1 Variables requises

Date

La variable de date contient les informations de temps et détermine l'unité de temps utilisée pour les prévisions.

1. Dans l'écran *Sélection des variables*, dans la section *Variables prédictibles conservées* (à gauche), sélectionnez la variable que vous souhaitez utiliser comme variable de date.
2. Cliquez sur le bouton > situé à gauche du champ *Date* dans la section *Variable requise* (en haut à droite). La variable est déplacée dans le champ *Date*.

> **i** Remarque
>
> Pour supprimer la variable de date, sélectionnez-la dans le champ *Date* puis cliquez sur le bouton > pour la ramener dans la section *Variables prédictibles conservées*.

Variable cible

La variable cible, ou signal, est celle que vous cherchez à expliquer ou dont vous voulez prédire les valeurs dans un jeu de données d'application.

- Pour choisir la variable cible :
  1. Dans l'écran *Sélection des variables*, dans la section *Variables prédictibles conservées* (à gauche), sélectionnez la variable que vous souhaitez utiliser comme variable cible.
  2. Cliquez sur le bouton > situé à gauche du champ *Cible* dans la section *Variable requise* (en haut à droite).
     La variable est déplacée dans le champ *Cible*.

     > **i** Remarque
     >
     > Pour supprimer la variable cible, sélectionnez-la dans le champ **Cible** puis cliquez sur le bouton < pour la ramener dans la section *Variables prédictibles conservées*.

5.4.5.1.2 Variable de poids

Une variable de poids vous permet d'attribuer un poids relatif à chaque observation qu'il décrit, et d'orienter activement le processus d'entraînement. Lorsque vous déclarez une variable comme variable de poids, cela signifie que chaque observation du jeu de données sera dupliquée proportionnellement à la valeur de cette variable.

- Pour choisir une variable de poids :
  1. Dans le champ *Variables prédictibles conservées*, sélectionnez la variable de poids.
  2. Cliquez sur le bouton > situé à droite du champ *Poids*.

     > **i** Remarque
     >
     > Pour supprimer la variable de poids, sélectionnez-la dans le champ *Poids*, puis cliquez sur le bouton <.

5.4.5.1.3 Affichage du signal

Vous pouvez afficher le signal contenu dans votre jeu de données sous forme de graphe en cliquant sur le bouton *Grapher les données* sous le champ *Signal*.

5.4.5.1.4 Variables exclues

Vous pouvez exclure des variables de la phase de modélisation. Exclure des attributs qui n'ont aucune influence sur les variables cible (par exemple, "numéro de compte") peut accélérer le processus de modélisation bien que les conserver ne nuise en rien à celui-ci. En revanche vous devez exclure les variables qui sont directement reliées à la variable cible, telles que des transformations de la variable cible ou des variables qui contiennent indirectement la variable cible. Par exemple, si la cible du modèle est la variable "a acheté le produit Oui/Non", vous devez exclure la variable "Montant de la facture" si elle contient le prix du produit.

- Pour exclure une variable :
  1. Dans le champ *Variables prédictibles conservées* (à gauche), sélectionnez les variables que vous souhaitez exclure de la modélisation.
     Remarque : Pour sélectionner toutes les variables d'un champ, cliquez à l'intérieur du champ, puis appuyez en même temps sur les touches `Ctrl` + `A`.
  2. Cliquez sur le bouton > près du coin supérieur droit du champ *Variables exclues*.

5.4.5.1.5 Réduction du jeu de données

Vous pouvez réduire la taille du jeu de données utilisé pour créer le modèle en définissant la dernière date du signal dans le premier champ de la section *Dernière date de l'entraînement*.

Le deuxième champ, indiquant le numéro de la ligne dans le jeu de données, est mis automatiquement à jour. Un message d'erreur s'affiche si la date n'existe pas dans le jeu de données.

Par défaut, la date correspondant à la dernière ligne de description du signal dans le jeu de données sert de dernière date de l'entraînement.

5.4.5.2 Sélection de variables

Ce panneau vous permet d'afficher le signal que vous avez sélectionné dans le champ *Date*.

- Pour afficher le signal
  1. Dans la liste *Date*, sélectionnez la variable contenant les informations temporelles.

     > **i** Remarque
     >
     > Par défaut, KTS utilise la première colonne du jeu de données comme variable de date.

  2. Dans la liste *Signal*, sélectionnez la variable contenant les informations de signal.
- Pour afficher une partie du signal
  1. Dans le champ *Première ligne*, saisissez le numéro de la ligne contenant le premier point du signal que vous voulez afficher.
  2. Dans le champ *Dernière ligne*, saisissez le numéro de la ligne contenant le dernier point du signal que vous voulez afficher.
  3. Cliquez sur le bouton *Actualiser* pour mettre à jour le graphe.

- Options disponibles

Le bouton...		vous permet de...
	Copier	copier les paramètres du graphique pour pouvoir les coller dans un tableur (comme Excel) et les utiliser pour générer un graphe.
	Imprimer	imprimer le graphe.
	Enregistrer	enregistrer le graphique sous forme d'image PNG.
	Copier l'image	copier l'image du graphe

5.4.5.3 Paramètres spécifiques du modèle

Ce panneau vous permet de définir les paramètres de modélisation avancés pour les séries temporelles.

- Général [page 133]
- Processus de modélisation disponibles [page 135]

5.4.5.3.1 Général

L'option *Ignorer les valeurs anormales pour le calcul de la tendance* a recours à une stratégie visant à réduire l'effet des valeurs anormales lors de l'estimation des régressions dans des tendances déterministes. Cette option entraîne une amélioration de l'entraînement de tendance.

L'option *Forcer un signal positif en sortie* permet de forcer la série temporelle à générer un modèle positif (avec des prévisions positives seulement).

Sélection des variables

Ce paramètre regroupe certains contrôles pour la sélection des variables. Lors de la sélection des variables, un processus de sélection automatique est appliqué sur les tendances ou les modèles autorégressifs pendant la compétition et le résultat n'est conservé que s'il améliore le modèle final.

Option	Description
Pourcentage de contributions variables à conserver	pourcentage des contributions qui sont conservées dans le processus de sélection automatique. La valeur par défaut est de 95 %.
Activer pour toutes les tendances extra-prédictibles	permet d'effectuer une sélection de variables sur toutes les tendances extra-prédictibles. Les variables de l'utilisateur sont conservées uniquement si elles apportent une contribution suffisante dans la régression de tendance. La case est cochée par défaut.
Activer pour tous les modèles auto-régressifs	permet d'effectuer pour tous les modèles auto-régressifs une sélection de variables automatique sur les valeurs antérieures du signal. Cette option conduit à un modèle d'auto-régression plus parcimonieux, c'est-à-dire un modèle plus simple et un niveau inférieur. L'option est désactivée par défaut.

Limitations sur les variables

La série temporelle génère automatiquement les variables nécessaires à la modélisation. Vous pouvez réduire le nombre de variables cycliques en modifiant le paramètre *Longueur maximale des cycles analysés* et celui des variables déphasées en modifiant le paramètre *Ordre maximal du modèle auto-régressif*.

Option	Description
Longueur maximale des cycles analysés	détermine la façon dont la série temporelle analyse les périodicités du signal. Il s'agit de la longueur du cycle le plus long que la série de temporelle va essayer de détecter. La valeur par défaut est de 450. Elle est également limitée par la taille du jeu de données d'entraînement. Vous pouvez désactiver l'analyse des variables cyclique en définissant ce paramètre sur zéro.
Ordre maximal du modèle auto-régressif	détermine la façon dont la série temporelle analyse les fluctuations aléatoires du signal. Ce paramètre définit la dépendance maximale du signal par rapport à ses propres valeurs antérieures. Vous pouvez lui attribuer la valeur "0" pour désactiver l'analyse des fluctuations.

> **i** Remarque
>
> La réduction du nombre de variables générées par la série temporelle permet de réduire la durée du calcul. Cependant, il est fortement recommandé de conserver les paramètres par défaut pour ne pas affecter la qualité de la modélisation.

5.4.5.3.2 Processus de modélisation disponibles

Pour modifier le processus de modélisation utilisé pour générer le modèle, vous pouvez sélectionner un des processus répertoriés dans le tableau suivant.

Processus de modélisation	Description
Défaut	Correspond à la modélisation standard par série temporelle Avec la modélisation standard par série temporelle, l'application essaie tous les modèles répertoriés dans le **Tableau 3**. Si aucun n'est utilisable, le processus par défaut est le *lissage exponentiel*.
Utiliser seulement les variables extra-prédictibles	Fonctionne comme modèle de classification/régression reposant sur les variables extra-prédictibles, avec le signal comme cible. Ce mode peut être utilisé pour affiner et valider les variables extra-prédictibles ou pour identifier celles qui sont inutiles.
Désactiver les tendances polynomiales	Génère tous les modèles sauf ceux contenant une tendance polynomiale.
Personnalisé	Permet d'activer ou désactiver les types de modèles générés par l'application lors de l'analyse du signal. Le tableau 3 récapitule les types de modèles qui peuvent être désactivés.
Lissage exponentiel	Génère les prévisions à l'aide du lissage exponentiel. Pour sélectionner la meilleure version de lissage exponentiel, l'application compare plusieurs processus de lissage : • double lissage, qui détecte la tendance ; • Triple lissage avec plusieurs tailles de cycle, à condition que le jeu de données d'entrainement contienne au moins deux cycles. L'application détermine les cycles d'après la granularité de la variable de date. Selon cette granularité, l'application teste plusieurs cycles sélectionnés auparavant. Par exemple, si la granularité de date est le mois, l'application teste deux cycles : un cycle de 3, qui correspond au trimestre et un cycle de 12 qui correspond à l'année. Le **Tableau 2** récapitule les cycles testés par l'application.
Régression linéaire	Génère les prévisions à l'aide d'une simple régression linéaire.

Cycles testés par l'application en fonction de la granularité de la variable temporelle

Granularité	Base temporelle	Longueur du cycle	Autres cycles testés
Seconde	Minute	60	360 (60*60) (heure)
Minute	Heure	60	1440 (60*24) (jour)
Heure	Jour	24	168 (semaine)
Jour	Semaine	7	30 (mois) - approximatif
Semaine	Année	52	12 (trimestre), 4 (mois) - approximatif
Mois	Année	12	3 (trimestre)
Trimestre	Année	4	
Semestre	Année	2	
Année	Décennie	10	

Le tableau suivant répertorie les types de modèles utilisés par le mode *Défaut*, qui peut être désactivé si nécessaire en utilisant le processus de modélisation *personnalisé*.

Types de modèles disponibles

Composant	Type de modèle	Description
Tendances	Décalage1	Valeur précédente du signal
	Décalage2	Valeur avant précédente
	Différenciation de second ordre	Tendance fondée sur la double différenciation pour propager la pente du signal
	Linéaire dans le temps	Régression linéaire dans le temps
	Polynomial dans le temps	Régression polynomiale dans le temps
	Linéaire sur extra-prédictibles	Régression linéaire sur les variables extra-prédictibles
	Linéaire dans le temps et linéaire sur extra-prédictibles	Régression linéaire dans le temps et sur les variables extra-prédictibles
	Polynomial dans le temps et linéaire sur extra-prédictibles	Régression polynomiale dans le temps et régression linéaire sur les variables extra-prédictibles
Périodicités	Cycliques	Détection des variables cycliques
	Saisonnalités	Détection des variables saisonnières
	Extra-prédictibles périodiques	Utilisation extra-prédictive comme périodiques
Fluctuations	Autorégressif	Modélisation auto-régressive

5.4.5.3.3 Modification du processus de modélisation

1. Dans la section *Modification du processus de modélisation*, cochez l'option souhaitée.
2. Si vous avez sélectionné l'option Personnalisé, décochez les types des modèles que vous souhaitez désactiver.
3. Cliquez sur le bouton *OK*. Le panneau *Récapitulatif des paramètres de modélisation* s'affiche.

5.4.6 Définition de paramètres pour les règles d'association

Cette section décrit les paramètres avancés spécifiques aux modèles de règles d'association.

5.4.6.1 Paramètres des règles d'association

Cet écran décrit les paramètres pour les fonctionnalités des règles d'association.

Sélection des colonnes

Dans la première partie de l'écran, sélectionnez les colonnes à utiliser pour réaliser la jointure entre le jeu de données de référence et le jeu de données d'événements.

1. Sélectionnez les identifiants à joindre en utilisant les listes déroulantes *Jeu de données de transaction* et *Jeu de données de référence*.
2. Sélectionnez la colonne contenant les éléments dans la liste déroulante *Colonne des éléments*.

Définition des paramètres

La seconde partie de l'écran décrit les paramètres qui seront utilisés pour calculer les règles d'association.

- Support minimal

Le *support d'une règle* est le nombre d'enregistrements qui vérifient l'antécédent de la règle. Dans le cas d'une règle du type X=>Y, le support est le nombre d'enregistrements contenant *le jeu d'éléments X et l'attribut Y*.

- Confiance minimale

La *confiance d'une règle* est le pourcentage d'enregistrements qui vérifient le conséquent de cette règle parmi ceux qui vérifient son antécédent. Dans le cas d'une règle de la forme X=>Y, la confiance de la règle est le nombre d'enregistrements contenant à la fois *le jeu d'éléments X et l'attribut Y* comparé au nombre d'enregistrements contenant *uniquement le jeu d'éléments X*.

- Longueur maximale

La *taille maximum d'une règle* est le nombre *total d'éléments* qu'elle contient, y compris l'antécédent et le conséquent. On appelle ça la *cardinalité* d'une règle.

5.4.6.2 Paramètres avancés des règles d'association

Paramètres avancés généraux

- Taille des blocs

Cette option vous permet de définir un nombre minimal de sessions pour lesquelles les règles d'association génèrent des règles temporaires.

Rappel : une session est un ensemble de transactions identifié par une clé unique. Par exemple, l'ensemble des achats effectués en une fois par un même client.

En usage standard (taille des blocs égale à 0), toutes les transactions en mémoire sont importées avant la génération des règles d'association correspondant aux paramètres du modèle (support minimal, confiance minimal, taille maximale des règles). Avec des jeux de données volumineux, cette méthode peut être très gourmande en temps et en espace mémoire. Pour gagner en vitesse et en espace mémoire, utilisez cette

option pour importer et générer les règles à partir de parties du fichier d'événements. À la fin de la phase d'apprentissage, ces règles sont rassemblées et les statistiques mises à jour pour l'ensemble du jeu de données.

Le nombre de règles trouvées peut être différent selon la méthode utilisée. Cependant, ces différences ne sont pas significatives si la taille des blocs est suffisante. La *taille des blocs* doit représenter environ 10 % du nombre total de sessions de votre jeu de données (à savoir 10 à 15 blocs selon la taille du jeu de données). Vous pouvez commencer avec 10 % et modifier cette valeur pour améliorer vos résultats. Si votre fichier de transactions contient moins de 5 000 lignes, vous devez utiliser une stratégie sans bloc.

Mais vous pouvez laisser Automated Analytics se charger de délimiter la taille des blocs grâce à l'option *Calculer automatiquement la taille des blocs*. La taille des blocs sera calculée de façon à en obtenir dix.

> **i** Remarque
>
> Les transactions présentes dans le jeu de données d'événement doivent être groupées par sessions. Cela signifie que le fichier d'événements doit être trié par sessions.

- Éviter les règles dérivées

Lorsque cette option est cochée, l'application ne génère que les règles primaires et non les règles dérivées. Par exemple, soit `R1 = X => Z` et `R2 X,Y =>Z` deux règles d'association, `R1` est appelée règle primaire et R2 règle dérivée.

Filtres sur les éléments conséquents

Avant de démarrer le processus d'apprentissage, vous pouvez sélectionner les éléments à afficher comme conséquents dans les règles générées en définissant soit les conséquents à conserver soit ceux à exclure de la modélisation. Si les deux listes contiennent des éléments, la liste des éléments à conserver sera utilisée pour la modélisation.

> **i** Remarque
>
> Avant de remplir une liste, vous devez connaître la *valeur exacte* des éléments que vous souhaitez inclure ou exclure.

- Pour sélectionner les éléments à conserver
  1. Sélectionnez l'onglet *Éléments à inclure dans le conséquent*. Cet onglet est sélectionné par défaut.
  2. Dans le champ de texte situé en bas du panneau, saisissez la valeur de l'élément que vous souhaitez voir apparaître dans les conséquents.
  3. Cliquez sur le bouton + situé à droite du champ de texte. L'élément apparaît dans la liste.
  4. Recommencez les étapes 2 et 3 pour tous les éléments que vous souhaitez inclure.
  5. Cliquez sur le bouton *OK* pour enregistrer la liste et retourner au panneau *Paramètres des règles d'association*.
- Pour sélectionner les éléments à exclure
  1. Sélectionnez l'onglet *Éléments à exclure du conséquent*.
  2. Dans le champ de texte situé en bas du panneau, saisissez la valeur de l'élément que vous ne souhaitez pas voir apparaître dans les conséquents.
  3. Cliquez sur le bouton + situé à droite du champ de texte. L'élément apparaît dans la liste.

4. Recommencez les étapes 2 et 3 pour tous les éléments que vous souhaitez exclure.
5. Cliquez sur le bouton *OK* pour enregistrer la liste et retourner au panneau *Paramètres des règles d'association*.

- Pour supprimer un élément de la liste
  1. Sélectionnez les éléments que vous voulez supprimer (utilisez la touche CTRL du clavier pour sélectionner plusieurs éléments).
  2. Cliquez avec le bouton droit de la souris dans la liste. Un menu contextuel s'affiche.
  3. Choisissez l'option *Supprimer les éléments sélectionnés*.
- Pour modifier un élément
  1. Sélectionnez l'élément que vous souhaitez modifier.
  2. Cliquez avec le bouton droit de la souris dans la liste. Un menu contextuel s'affiche.
  3. Choisissez l'option *Modifier l'élément sélectionné*. La valeur de l'élément s'affiche dans le champ de texte.
  4. Modifiez la valeur.
  5. Cliquez sur le bouton + pour valider la modification.
- Trier une liste alphabétiquement
  - Cochez l'option *Tri alphabétique* pour voir les conséquents trié par ordre alphabétique.

5.4.7 Définition des paramètres pour la journalisation d'événements

Cette section décrit les paramètres avancés spécifiques aux modèles utilisant la journalisation d'événements.

5.4.7.1 Paramètres de journalisation des événements

Ce panneau vous permet de définir les paramètres suivants :

- La jointure entre les données des événements et les données de référence
- Le mode de stockage des colonnes temporaires lors de la génération du modèle
- Le préfixe des variables générées par la *journalisation des événements*.

- Jointure des jeux de données [page 139]
- Sélection du type de stockage [page 140]
- Sélection du préfixe des variables [page 140]

5.4.7.1.1 Jointure des jeux de données

1. Dans la première partie de l'écran, *Colonnes de jointure* décrit les colonnes à utiliser pour réaliser la jointure entre le *jeu de données de référence* et le jeu de données des événements (*jeu de données de journal*) : sélectionnez les identifiants à joindre à l'aide des menus déroulants. Seules des variables nominales

peuvent être utilisées comme colonnes de jointure. Si la variable que voulez utiliser ne figure pas dans la liste, revenez au panneau *Décrire les données* et sélectionnez le type *Nominale*.

2. Sélectionnez la variable date des données des événements dans le menu déroulant *Colonne date du Log*.

> **i** Remarque
>
> Reportez-vous au document *Data Modeling Specifications* pour connaître les formats de date autorisés.

5.4.7.1.2 Sélection du type de stockage

Lorsque vous créez un modèle, la journalisation des événements génère une grande quantité de colonnes temporaires. Vous pouvez indiquer si les données générées doivent être stockées dans l'espace mémoire ou sur disque.

Les deux modes disponibles sont en *Mémoire* ou *Sur disque* :

- *Mémoire* signifie que les calculs seront stockés dans la mémoire interne. Ce mode est plus rapide mais son utilisation limite la quantité de données pouvant être traitées. Lorsque vous utilisez cette option, vous devez faire attention de fournir à la journalisation des événements une quantité de données traitable pour éviter tout plantage de l'application.
- *Sur disque* (sélectionné par défaut) signifie que les calculs internes seront stockés dans un fichier sur le disque dur. Dans ce but, la journalisation des événements traitera les données par blocs. Étant donnée une limite de mémoire, elle calculera combien de données peuvent être stockées en mémoire, effectuera les calculs, les enregistrera sur le disque, et ainsi de suite. Ce mode est plus lent que le mode *Mémoire* car il nécessite un passage sur le jeu de données d'événements pour chaque bloc. Cependant, vous pouvez personnaliser cette fonctionnalité en spécifiant la taille limite de la mémoire à utiliser. Indiquez pour cela dans le fichier de configuration une valeur numérique pour le paramètre *KELMemoryThreshold* représentant l'espace mémoire en méga-octets que la journalisation des événements peut consommer. Si vous attribuez par exemple la valeur 1024 à KELMemoryThreshold, la journalisation des événements n'utilisera au maximum que 1024 Mo de mémoire en mode *Sur disque*. La valeur par défaut de ce paramètre est 512 Mo.
  Notez que cette valeur est une estimation et n'est applicable qu'à la journalisation des événements. Autrement dit, la quantité de mémoire utilisée par ce module peut être légèrement supérieure ou très inférieure à ce chiffre (si tous les calculs peuvent être stockés en mémoire par exemple). Cependant la taille globale du processus de l'application peut être supérieure à ce chiffre car d'autres fonctionnalités peuvent intervenir dans le processus d'apprentissage. L'utilisation d'une interface graphique, par exemple, consomme également de la mémoire. Ce point n'est pas pris en compte par cette option.

5.4.7.1.3 Sélection du préfixe des variables

Cette option vous permet de choisir le préfixe qui identifiera les variables créées par la *journalisation des événements*. Par défaut, la valeur du préfixe est *el*.

5.4.7.2 Sélection des variables de journalisation des événements pour les fonctions

L'écran *Journalisation des événements - Sélection des variables pour les fonctions* vous permet de spécifier les agrégats que vous voulez calculer sur vos données de transactions (ou d'événements). Cet écran est divisé en trois parties :

1. Tout d'abord la définition des agrégats
2. Puis la définition des périodes sur lesquelles les agrégats doivent être calculés
3. Enfin la définition des filtres ou pivots facultatifs

Deux types de configuration sont possibles :

-
-

5.4.7.2.1 Configuration simple

Définition des agrégats

- Pour définir de nouveaux agrégats :
  1. Cliquez sur *Agrégats* pour afficher l'onglet correspondant.

     > **i** Remarque
     >
     > Cette section est affichée par défaut à l'ouverture du panneau *Journalisation des événements - Sélection des variables pour les fonctions.*

  2. Vous pouvez appliquer des méta-opérateurs à l'opérateur *Count*. Cet opérateur, qui est toujours appliqué, compte le nombre d'événements dans chaque période définie (pour plus de détails sur les méta-opérateurs, reportez-vous au tableau ci-dessous).
  3. Dans le tableau, cochez la case **Agrégats sélectionnés** correspondant à la variable et à l'opérateur à utiliser pour créer l'agrégat. Quatre types d'opérateurs sont proposés pour chaque variable : *Max*, *Sum*, *Average* et *Min*.
     Selon le résultat souhaité, vous pouvez appliquer des *méta-opérateurs* sur les agrégats que vous créez. Les méta-opérateurs peuvent être appliqués à un même agrégat.

Le méta-opérateur...	ajoute à l'agrégat...	calcule...
Variation	une colonne pour chaque ensemble de deux périodes consécutives (c'est-à-dire n-1)	la différence entre les valeurs de deux périodes consécutives pour l'ensemble des périodes de l'agrégat
Variation en pourcentage	une colonne pour chaque ensemble de deux périodes consécutives	la différence en pourcentage entre les valeurs de deux périodes consécutives pour l'ensemble des périodes de l'agrégat

Le méta-opérateur...	ajoute à l'agrégat...	calcule...
Accumulation	une colonne par période, sauf pour la première période	l'accumulation totale actuelle pour chaque période de l'agrégat
Accumulation à rebours	une colonne par période, sauf pour la dernière période	l'accumulation totale actuelle pour chaque période de l'agrégat, calculée à rebours
Somme globale	une colonne	la somme des valeurs de toutes les périodes de l'agrégat

4. Par défaut, toutes les variables numériques sont listées dans le tableau, cependant vous pouvez choisir de n'afficher que les variables continues en cochant la case *N'afficher que les variables continues*.

Paramétrage des dates

- Pour définir les périodes :
  1. Cliquez sur *Périodes* pour afficher l'onglet correspondant.
  2. Dans le cadre *Date de référence*, paramétrez la date de référence comme expliqué dans la section **Date de référence** ci-dessous.
  3. Dans le cadre *Définir une période puis l'ajouter à la liste*, définissez une ou plusieurs périodes comme expliqué dans la section **Paramétrage des périodes** ci-dessous.

Date de référence

La *date de référence* est la date à partir de laquelle vont être définies les périodes à analyser. Cette date peut être :

- soit *fixe*, c'est à dire une date spécifique, par exemple 24-05-2004,
- soit *variable*, c'est à dire une date définie par une des variables du jeu de données. Par exemple, la date du premier achat, la date de résiliation d'un abonnement, etc.

- Pour définir la date de référence
  1. Sélectionnez le type de la date de référence en cochant l'une des options *Fixe* ou *Variable*.
  2. Selon l'option choisie ci-dessus, spécifiez une date fixe dans le champ de texte correspondant, ou sélectionnez une des variables date de la table de référence proposées dans la liste.

Paramétrage des périodes

Deux types de périodes existent :

- les *périodes simples*, qui sont des périodes successives de longueur définie ;
- les *périodes cycliques*, qui sont des périodes récurrentes.

> **i** Remarque
>
> - Notez que la numérotation des colonnes de sortie est différente pour les périodes simples et les périodes cycliques. Supposons, par exemple, que nous ayons la date fixe du 18 août 2007 et que notre demande porte sur les 24 derniers mois (période simple). Les colonnes de sortie seront alors numérotées par mois de 0 à 23, le mois 23 comportant les entrées entre le 19 juillet et le 18 août 2007.

- Si nous spécifions une période cyclique par mois, les colonnes de sortie seront numérotées de 0 (janvier) à 11 (décembre), indépendamment de la date fixe.

Périodes simples

L'interface vous permettant de paramétrer les périodes est écrite en langage naturel. Chaque paramètre pouvant être modifié est signalé par un *lien hypertexte* (en *gras* et *souligné*).

- Pour définir une période simple
  1. Cochez l'option *Période simple* dans le cadre *Définir une période*.
  2. Dans la phrase affichée, cliquez sur le mot ou le nombre correspondant au paramètre à modifier. Une période est définie par :
     - un nombre de périodes,
     - une durée,
     - une date de début, qui peut être antérieure ou postérieure à la date de référence choisie. La date de début est incluse dans la période définie.
  3. Répéter cette opération pour tous les paramètres à définir.

Périodes cycliques

- Pour définir des périodes cycliques :
  1. Cochez l'option *Période cyclique* dans le cadre *Définir une période*.
  2. Dans la liste déroulante *Créer une période pour chaque*, sélectionnez le cycle.
  3. Vous pouvez appliquer ces périodes sur l'intégralité du jeu de données ou sélectionner une ou plusieurs périodes successives.
     - Sélectionnez l'option *Tous les événements* pour appliquer les périodes cycliques à l'ensemble du jeu de données.
     - Sélectionnez l'option *Uniquement les événements compris dans la plage temporelle suivante* pour limiter la période cyclique à une plage temporelle précise.

Filtrage du jeux de données

Vous pouvez filtrer votre jeu de données en spécifiant les lignes à filtrer en fonction des valeurs d'une variable.

- Pour filtrer le jeu de données
  1. Ouvrez l'onglet *Filtres (optionnel)*.
  2. Selon le type de filtre à créer, sélectionnez soit l'option *Filtre*, soit l'option *Pivot*.
     - L'option *Filtre* vous permet de filtrer vos données selon les valeurs d'une variable.
     - L'option *Pivot* vous permet de créer une variable de journalisation des événements pour chaque valeur sélectionnée.
  3. Sélectionnez le *type de filtre* à appliquer en cochant l'une des options disponibles :

L'option...	vous permet de filtrer en...
Valeurs exclues	excluant les enregistrements pour lesquels la variable sélectionnée a les valeurs indiquées.

L'option...	vous permet de filtrer en...
Valeurs conservées	conservant uniquement les enregistrements pour lesquels la variable sélectionnée a les valeurs indiquées.

4. Dans la liste déroulante *Variables*, sélectionnez la variable sur laquelle vous voulez filtrer.
5. Pour ajouter des catégories à la table, vous pouvez :
   - extraire automatiquement les catégories des variables en cliquant sur les jumelles ( ) situées à côté de la liste, puis sélectionner les valeurs en cochant la case *Sélection* correspondante.
   - ou saisir une valeur dans le champ *Nouvelle catégorie* et cliquez sur le bouton +.

   > **i** Remarque
   >
   > Le nombre de variables créées est indiqué en bas du panneau. Ce nombre augmente de façon exponentielle en cas de filtrage par pivot. Plus le nombre de variables est élevé, plus l'apprentissage est long.
6. Cochez la case *Créer aussi des agrégats non filtrés* pour créer une variable agrégée non filtrée.

Vous pouvez avoir besoin de supprimer une valeur que vous avez déjà définie dans le filtre. Dans ce cas :

1. Sélectionnez la valeur dans la liste sous le champ *Valeur(s)*.
2. Cliquez sur le bouton -. La valeur sélectionnée sera supprimée.

5.4.7.2.2 Configuration avancée

La *configuration avancée* vous permet de créer plusieurs ensembles d'agrégats, plusieurs périodes et plusieurs filtres et de combiner ces trois éléments comme bon vous semble.

Définition des ensembles d'agrégats

- Pour définir les ensembles d'agrégats :

Pour chaque ensemble d'agrégats à créer :

1. Suivez la procédure *Définition des agrégats* décrite dans la *configuration simple*.
2. Dans le champ *Nom de l'ensemble d'agrégats*, saisissez un nom pour l'ensemble à créer.
3. Cliquez sur le bouton + pour ajouter l'ensemble d'agrégats à la *Liste des ensembles d'agrégats*.

- Pour supprimer un ensemble d'agrégats :
  1. Dans la *Liste des ensembles d'agrégats*, sélectionnez l'ensemble à supprimer.
  2. Cliquez sur le bouton - situé sous la liste. L'ensemble d'agrégats sélectionné est supprimé.

Définition des périodes

- Pour définir les périodes :

Pour chaque période à créer :

1. Suivez la procédure décrite dans la section **Configuration simple** ci-dessus.
2. Dans le champ *Nom de la période*, saisissez un nom pour la période à créer.
3. Cliquez sur le bouton + pour ajouter la période à la *Liste des périodes*.

- Pour supprimer une période :
  1. Dans la *Liste des périodes*, sélectionnez la période à supprimer.
  2. Cliquez sur le bouton -. La période sélectionnée est supprimée.

Définition des filtres et pivots

- Pour définir les filtres ou pivots :

Pour chaque filtre ou pivot à créer :

1. Suivez la procédure décrite dans la section **Configuration simple** ci-dessus.
2. Dans le champ *Nom du filtre ou du pivot*, saisissez un nom pour le filtre ou le pivot à créer.
3. Cliquez sur le bouton + pour ajouter le filtre à la *Liste des filtres et pivots*.
4. La case à cocher *Ajouter aussi les agrégats non filtrés* vous permet de créer un filtre vide, c'est-à-dire un filtre qui fournira toutes les données.

- Pour supprimer un filtre ou un pivot :
  1. Dans la *Liste des filtres et pivots*, sélectionnez le filtre ou pivot à supprimer.
  2. Cliquez sur le bouton -. Le filtre ou le pivot sélectionné est supprimé.

5.4.8 Définition de paramètres pour l'analyse de séquences

5.4.8.1 Paramètres de l'analyseur de séquences

L'écran *Paramètres de l'analyse de séquences* vous permet de définir certains paramètres à l'aide des tâches suivantes :

- Joindre vos données de référence à vos données de transactions
- Calculer les séquences intermédiaires

- Joindre vos données [page 146]
- Séquences intermédiaires [page 146]

- Filtrage des événements par période [page 146]

5.4.8.1.1 Joindre vos données

Pour agréger les données de référence avec les données des événements, vous devez joindre les deux tables et indiquer la colonne de chaque table qui correspond à l'identifiant de référence.

Dans les champs *Colonnes de jointure*, sélectionnez les variables correspondant à l'identifiant du client dans chacun des jeux de données. L'information contenue dans les deux variables sélectionnées doit être identique.

Dans le champ *Colonne date du Log*, sélectionnez la variable correspondant à la date et/ou l'heure des données du log.

5.4.8.1.2 Séquences intermédiaires

Le mode *Séquences Intermédiaires* vous permet d'obtenir des informations supplémentaires sur les transitions et séquences existant dans vos jeux de données :

- ordre des étapes
- détails des étapes
- continuité de la session pour chaque étape

5.4.8.1.3 Filtrage des événements par période

La section *Fenêtre de temps* vous permet de filtrer les événements utilisés pour créer le modèle en définissant une période à l'aide de dates fixes ou de valeurs existant dans le jeu de données. Vous disposez des options suivantes pour filtrer le jeu de données des événements :

Option	Description
Infini	Aucune période n'est définie : tous les événements seront définis.
Fixe	Seuls les événements pour lesquels la valeur de la *Colonne date du log* se situe entre les deux dates sélectionnées seront conservés.
Entre deux colonnes de date	Seuls les événements pour lesquels la valeur de la *Colonne date du log* se situe entre les valeurs des deux colonnes de dates sélectionnées seront conservés. Par exemple, vous pouvez sélectionner les colonnes de dates correspondant au début et à la fin d'une période d'essai gratuite, ces dates pouvant être différentes pour chaque client.

Option	Description
Relatif à une colonne de date	Seuls les événements pour lesquels la valeur de la *Colonne date du log* se trouve dans l'intervalle défini relativement à la colonne de date sélectionnée seront conservés. Par exemple, vous pouvez utiliser la date d'achat d'une carte de crédit comme date de référence et sélectionner tous les événements qui ont eu lieu durant les trois mois précédant cette date.

> **i** Remarque
>
> - Choisissez soigneusement la période ; celle-ci doit comporter des événements existant dans le jeu de données ou vous risquez d'obtenir des résultats aberrants (KI négatif, KR égal à 1, etc.).
> - Veillez à exclure les dates de début et de fin de la période définie.

- Pour utiliser tous les événements du jeu de données
  - Cochez l'option *Infini*.
- Pour utiliser uniquement les événements se produisant durant une période fixe
  1. Cochez l'option *Fixe*.
  2. Dans le champ *Début*, sélectionnez la date avant laquelle aucun événement ne doit être conservé.
  3. Dans le champ *Fin*, sélectionnez la date après laquelle aucun événement ne doit être conservé.
- Pour utiliser uniquement les événements se produisant entre deux colonnes de date
  1. Cochez l'option *Entre deux colonnes de date*.
  2. Dans le champ *Début*, sélectionnez la colonne de date qui contient la date avant laquelle aucun événement ne doit être conservé.
  3. Dans le champ *Fin*, sélectionnez la colonne de date qui contient la date après laquelle aucun événement ne doit être conservé.
- Pour utiliser uniquement les événements se produisant dans un intervalle relatif à une colonne de date
  1. Cochez l'option *Relatif à une colonne de date*.
  2. Dans la liste *Date*, sélectionnez la colonne qui contient la date à utiliser comme référence pour la période.
  3. Dans le champ *Entre*, saisissez le nombre d'unité qui indiquera quand commence la période. Le tableau ci-dessous résume les valeurs possibles pour définir le début de la période.

Valeur	Signification
entier négatif	la période commence avant la date de référence
0	la période commence à la date de référence
entier positif	la période commence après la date de référence

  4. Dans le champ *et*, saisissez le nombre d'unité qui indiquera quand se termine la période.
  5. Dans la dernière liste déroulante, sélectionnez l'unité à utiliser pour définir la période.
     Par exemple, si vous saisissez les paramètres *Date* AchatCarte *Entre* -3 **et** 0 Mois, seuls les événements se produisant dans les trois mois précédant la date d'achat de la carte seront conservés pour chaque client.

5.4.8.2 Sélection des variables d'analyseur de séquences pour les fonctions

L'écran *Sélection des variables pour les fonctions d'analyse de séquences* vous permet de spécifier le type de statistiques que vous voulez calculer sur vos données de transactions ou d'événements.

Définition des opérations

Plusieurs colonnes sont créées en standard par KSC pour chaque identifiant du fichier de référence. Lorsque un identifiant n'a pas de transactions associées, ces colonnes ont une valeur nulle.

KSC_Start_Date : Horodatage de la première transaction dans le log pour chaque identifiant de référence

KSC_End_Date : Horodatage de la dernière transaction dans le log pour chaque identifiant de référence

KSC_TotalTime : Temps écoulé en secondes entre *KSC_Start_Date* et *KSC_End_Date*

KSC_Number_Events : Nombre de transactions dans le log associées à chaque identifiant de référence

En plus des colonnes standard de l'analyseur de séquences, trois types d'opérations sont disponibles :

- *Compter*,

Quand vous sélectionnez l'option *Compter*, l'analyseur de séquence crée une colonne pour chaque valeur de la variable sélectionnée.

L'option *Compter* code la séquence en créant une colonne par catégories valides de la variable spécifiée. Chaque catégorie valide est appelée "état". Les catégories qui n'apparaissent qu'une seule fois pour les transactions associées à un identifiant du fichier référence présent dans le jeu de données d'entraînement sont ignorées.

- *Compter les transitions*,

Quand vous choisissez l'option *Compter les transitions*, l'analyseur de séquences crée une colonne pour chaque transition entre catégories trouvée dans le jeu de données sélectionné

Compter les transitions code les séquences en créant une colonne par transition entre deux catégories de la colonne spécifiée. Chaque transition valide est appelée "transition d'état". Les transitions d'état qui n'apparaissent qu'une seule fois pour les transactions associées à l'identifiant de référence présent dans le jeu de données d'*entraînement* sont ignorées. Une colonne *KxOther* distincte est créée pour les transitions rares, sur la base du seuil défini par le curseur Filtre, de la même façon qu'une colonne *KxOther* est créée pour les décomptes.

- *Premier et dernier*.

L'option *Premier et dernier* crée deux colonnes, appelées "*FirstState*" et "*LastState*", contenant les catégories de la variable choisie correspondant à la première et à la dernière transactions du journal pour chaque identifiant de référence. Les colonnes "*FirstState*" et "*LastState*" sont créées automatiquement lorsque l'une des options *Compter* et *Compter les transitions* est sélectionnée.

5.4.8.3 Paramètres avancés de l'analyseur de séquences

Cet écran vous permet de définir les paramètres avancés de l'analyseur de séquences fourni par Data Manager.

Stockage des informations de comptage

Lorsque vous créez un modèle, l'analyseur de séquences génère une grande quantité de colonnes temporaires. Vous pouvez indiquer si les données générées doivent être stockées dans l'espace mémoire ou sur disque.

L'option *En mémoire* est sélectionnée par défaut.

Préfixe des variables générées par l'analyseur de séquences

Vous pouvez définir un préfixe spécifique qui identifiera les variables créées par Data Manager. Par défaut, la valeur du préfixe est ksc.

Filtrage des événements

L'option *Filtrage* permet de regrouper des catégories rares en une seule appelée *KxOther*. Il arrive fréquemment que les fichiers de transactions comportent un grand nombre de catégories peu fréquentes qui ne constituent pas en elles-mêmes des prédicateurs fiables. Regrouper ces catégories rares en une seule apporte souvent un avantage prédictif. La règle Filtrage vous permet de choisir les catégories à conserver dans des colonnes distinctes sur la base d'un pourcentage de la totalité du fichier de transactions. Les catégories correspondant au pourcentage de transactions restant sont regroupées dans la colonne *KxOther* automatiquement générée par l'analyseur de séquences.

Si, par exemple, vous déplacez le curseur *Filtrage* sur 90 %, le nombre total de transactions obtenu en additionnant toutes les catégories affectées à des colonnes distinctes ne doit pas dépasser 90 % du nombre total de transactions. Les catégories formant les 10 % restants sont regroupées dans *KxOther*.

Vous pouvez également définir un *seuil* de façon à ce que les transitions dont la durée entre deux événements est supérieure à ce seuil soient ignorées dans le comptage des transitions.

- Pour définir un seuil
  1. Cochez la case *Filtrer les transitions supérieures à*.
  2. Dans le champ Nombre, saisissez le nombre d'unités qui définira le seuil.

Dans la liste déroulante, sélectionnez l'unité utilisée pour définir le seuil.

Vous pouvez également définir un seuil de façon que le nombre de transitions ne soit mis à jour que si le temps écoulé durant la transition entre deux événements est inférieur au seuil défini.

Vous pouvez également définir un seuil de façon que les transitions dont la durée entre deux événements est supérieure à ce seuil soient ignorées dans le comptage des transitions.

- Pour définir un seuil
  1. Cochez la case *Filtrer les transitions supérieures à*.
  2. Dans le champ Nombre, saisissez le nombre d'unités qui définira le seuil.
  3. Dans la liste déroulante, sélectionnez l'unité utilisée pour définir le seuil.

5.4.9 Définition de paramètres pour l'analyse de texte

Cette section décrit les paramètres avancés spécifiques aux modèles utilisant l'analyse de texte.

5.4.9.1 Paramètres de l'encodage textuel

Ce panneau vous permet de définir les paramètres de langue.

Référentiel de définitions des langues

Le référentiel de définitions des langues contient tous les fichiers nécessaires pour définir et déterminer une langue, comme par exemple la liste des mots vides et les règles de lemmatisation.

Dans la liste *Dossier*, sélectionnez le dossier contenant le fichier de données linguistiques.

Langues prises en charge/Langues exclues

Une fois que le dossier a été sélectionné, la liste des langues prises en charge est actualisée.

Vous pouvez exclure une langue en la sélectionnant dans la liste *Langues prises en charge*, puis en cliquant sur le bouton > à gauche de la liste *Langues exclues*.

Vous pouvez rajouter une langue qui a été exclue en la sélectionnant dans la liste *Langues exclues*, puis en cliquant sur le bouton < à gauche de la liste *Langues exclues*.

Pour visualiser le contenu des fichiers de données linguistiques d'une ou plusieurs langues, sélectionnez les langues choisies dans la liste *Langues prises en charge*, puis cliquez sur la *loupe* située juste en-dessous de la liste. Une nouvelle fenêtre s'ouvre dans laquelle chaque onglet correspond à un fichier de données.

5.4.9.2 Encodage textuel - Paramètres avancées des langues

Ce panneau vous permet de définir votre propre liste de mots vides indépendamment de la langue des jeux de données utilisés.

> **i** Remarque
>
> Les mots vides sont des mots communs qui n'apportent aucune information, tels que des mots de liaison, des articles, etc. Il peut également s'agir de mots qui apparaissent trop fréquemment dans le jeu de données pour fournir des informations utiles.

- Pour ajouter un mot vide à la liste
  1. Dans le champ de texte situé en bas du panneau, saisissez le mot à ajouter à la liste de mots vides.
  2. Cliquez sur le bouton +. Le mot apparaît dans la liste au-dessus.
- Pour supprimer un mot vide de la liste
  1. Sélectionnez dans la liste le mot à supprimer.
  2. Cliquez sur le bouton - situé à droite du champ de texte. Le mot disparaît de la liste.
- Pour enregistrer la liste de mots vides
  1. Créer la liste en ajoutant les mots comme indiqué ci-dessus.
  2. Cliquez sur le bouton *Enregistrer*.
  3. Dans le champ *Dossier*, sélectionnez le dossier dans lequel vous voulez enregistrer la liste des mots vides.
  4. Dans le champ *Liste*, saisissez le nom du fichier dans lequel sera enregistrée la liste.
  5. Cliquez sur le bouton *OK*. La liste est enregistrée.
- Ouvrir une liste de mots vides
  1. Cliquez sur le bouton *Ouvrir*. Une fenêtre s'ouvre.
  2. Utilisez le bouton *Parcourir* du champ *Dossier* pour sélectionner le dossier où se trouve la liste à ouvrir.
  3. Utilisez le bouton *Parcourir* du champ *Liste* pour sélectionner le fichier contenant la liste à ouvrir.
  4. Cliquez sur le bouton *OK*. La liste est chargée.
- Pour trier la liste par ordre alphabétique
  - Cochez la case *Tri alphabétique*.

5.4.9.3 Paramètres de l'encodage textuel

- Paramètres de construction du dictionnaire [page 152]
- Paramètres de codage [page 153]

5.4.9.3.1 Paramètres de construction du dictionnaire

Le dictionnaire est formé de racines, c'est-à-dire des mots ou des termes significatifs. Vous pouvez définir les paramètres de construction du dictionnaire suivants :

- *Suppression des mots vides*
  Cochez cette option pour supprimer les mots vides de la liste des racines.
- *Réduction par règle de lemmatisation*
  Cette option vous permet de supprimer les affixes pour limiter le nombre de racines.
- *Regroupement de concepts*
  Cette option vous permet d'utiliser un fichier externe associant des termes (c'est-à-dire des groupes de mots désignant un concept unique, tels que "la Maison Blanche" ou "carte bancaire") à des concepts. Étant donné que cette option traite des groupes de mots, elle est appliquée avant la suppression des mots vides et la règle de lemmatisation. Vous pouvez créer votre propre dictionnaire de concepts en créant un fichier texte nommé ConceptList_<CodeLangue> (sans extension) et contenant à chaque ligne un terme et le concept qui lui est associé. Par exemple, vous pouvez regrouper les concepts dans le domaine des compagnies aériennes :
  ```
  word=concept
  business-class=BusinessClass
  first-class=FirstClass
  flying-blue=FlyingBlue
  ```
  Ou vous pouvez créer le concept "cartebancaire" qui remplacera toutes les cartes bancaires (telles que "Carte Visa", "American Express", "Carte bleue") :
  ```
  credit-card=creditcard
  american-express=creditcard
  visa-card=creditcard
  mastercard=creditcard
  ```
  Notez que vous devez remplacer les espaces entre les termes par des `dashes`, séparer les termes et concepts par le signe égal `"="` et écrire les termes en **minuscules** (car les majuscules sont retirées avant le regroupement de concepts).
  Et notez qu'il faut préciser le terme au singulier et au pluriel pour prendre en compte toutes les occurrences.
  La liste de concepts est dépendante de la langue ; la liste appropriée est automatiquement sélectionnée une fois que la langue du jeu de données a été identifiée ou choisie par l'utilisateur.
- *Remplacement des synonymes*
  Cette option vous permet d'utiliser un fichier externe définissant des racines synonymes. Il sera utilisé pour remplacer certaines racines par une seule spécifiée par l'utilisateur. Cette option est appliquée après la suppression des mots vides et la règle de lemmatisation. Vous pouvez créer votre propre dictionnaire de synonymes en créant un fichier texte nommé *SynonymList_<CodeLangue>* et contenant sur chaque ligne une racine trouvée par encodage textuel associée à la racine synonyme, comme indiqué ci-dessous :
  ```
  <found root>=<replacement root>
  ```

  > **i** Remarque
  >
  > La racine et son synonyme associé sont séparés par le signe égal (=).

  La liste de synonymes est dépendante de la langue ; la liste appropriée est automatiquement sélectionnée une fois que la langue du jeu de données a été identifiée ou choisie par l'utilisateur.
- *Nombre maximal de racines générées*

Cette option vous permet de choisir combien de racines vous souhaitez conserver dans le dictionnaire. Par défaut, les racines les plus fréquentes sont conservées, mais vous pouvez aussi choisir d'ignorer une partie des racines les plus fréquentes en cliquant sur le bouton *Avancé*

5.4.9.3.2 Paramètres de codage

Chaque racine est transformée en une colonne et lorsque la racine apparaît dans un texte, sa présence peut être codée selon une des méthodes suivantes :

- *Booléen* : si le mot est présent dans le texte, la colonne prend la valeur 1 et sinon la valeur 0.
- *Fréquence (TF)* : nombre de fois où la racine apparaît dans le texte en cours divisé par le nombre total de mots du document.
- *TF-IDF* : mesure de l'importance générale d'une racine dans le document en cours par rapport à l'ensemble des documents sur la base de la fréquence.
  `TFIDF = TF*log10(TotalNumberOfDocuments/NumberOfDocumentsContainingTheRoot)`
- *Compte (TC)* : nombre de fois où la racine apparaît dans le texte en cours.
- *TC-IDF* : mesure de l'importance générale d'une racine dans le document en cours par rapport à l'ensemble des documents sur la base du compte.
  `TCIDF = TC*log10(TotalNumberOfDocuments/NumberOfDocumentsContainingTheRoot)`

5.4.9.4 Encodage textuel - Paramètres avancées

Ce panneau vous permet de choisir un pourcentage de racines à exclure. Vous pouvez modifier les valeurs en faisant glisser les curseurs. Par défaut, 5 % des racines les moins fréquentes sont ignorées. Mais il peut être s'avérer intéressant d'exclure certaines des racines les plus fréquentes car elles peuvent être comparées à des mots vides et, à ce titre, n'apportent aucune information nouvelle.

Une fois le pourcentage souhaité indiqué, cliquez sur le bouton *Valider*.

5.5 Apprentissage d'un modèle

5.5.1 Apprentissage du modèle

Ce panneau affiche le processus d'apprentissage du modèle.

- Aperçu [page 154]
- Notifications [page 155]
- Classification/Régression et segmentation [page 155]
- Série temporelle [page 158]

- Règles d'association [page 160]
- Journalisation des événements et analyseur de séquences [page 160]
- Social [page 161]
- Recommandation [page 162]

5.5.1.1 Aperçu

Aperçu

Date de création	Date et heure de la création du modèle
Temps d'apprentissage	Temps total pour l'apprentissage du modèle
Nom du moteur	Selon le composant utilisé : • Kxen.RobustRegression (classification/régression) • Kxen.SmartSegmenter (segmentation/mise en cluster) • Kxen.TimeSeries (série temporelle) • Kxen.AssociationRules (règles d'association) • Kxen.EventLog (journalisation des événements) • Kxen.SequenceCoder (analyseur de séquences) • Kxen.SocialNetworks (analyse des réseaux sociaux)
Auteur	Nom de l'utilisateur

Détails

Éléments	Nombre d'éléments correspondant aux valeurs définies lors du paramétrage du recommandeur
Règles	Nombre de règles trouvées par ce recommandeur

Transactions

Nb. d'utilisateurs distincts	Nombre d'utilisateurs distincts identifiés dans le jeu de données d'entraînement
Nb. d'éléments distincts	Nombre d'utilisateurs distincts identifiés dans le jeu de données d'entraînement
Nb. de paires utilisateur/élément	Nombre de paires distinctes reliant un utilisateur et un élément
Nombre de lignes utilisées	Nombre de lignes du jeu de données une fois le filtre appliqué
Meilleures ventes	Nombre d'éléments identifiés comme "meilleures ventes", c'est-à-dire qui apparaissent dans les transactions un nombre de fois supérieur au seuil défini dans la section "Identifier les meilleures ventes" lors du paramétrage du modèle.

5.5.1.2 Notifications

Variables Monotones Détectées	Indique si des variables monotones ont été trouvées dans le jeu de données, c'est-à-dire des variables dont le sens de variation est constant, dans l'ordre de lecture des données dans le jeu de données d'entraînement.
Variables Suspectes Détectées	Ce rapport présente une liste de variables considérées comme suspectes. Ces variables suspectes ont une capacité prédictive > 0,9 et sont très fortement corrélées à la variable cible. Cela signifie que ces variables apportent probablement une information biaisée et qu'elles ne devraient pas être utilisées pour la modélisation. Vous devez porter une attention particulière à ces variables. Un rapport fournit la liste précise des variables suspectes avec leur valeur (voir Rapports statistiques > Analyse avancée > Variables suspectes).

5.5.1.3 Classification/Régression et segmentation

Pour chaque variable nominale :

<Nom>	Nom de la variable cible
Clé cible	Valeur attendue de la clé cible
Fréquence des catégories cible	Pourcentage d'observations de toutes les valeurs cible dans le jeu de données d'entraînement dans le cas d'une cible nominale

Variables cible continues

Pour chaque variable cible continue :

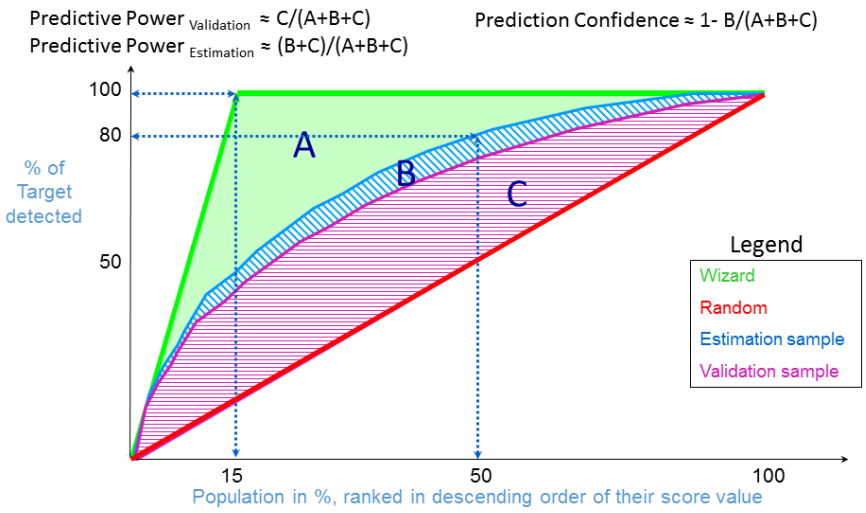
<nom de la cible>	Nom de la variable cible
Min	Valeur minimale de la variable cible dans le jeu de données d'entraînement
Max	Valeur maximale de la variable cible dans le jeu de données d'entraînement
Moyenne	Moyenne des valeurs de la variable cible pour le jeu de données d'entraînement
Écart-type	Moyenne de la distance entre les valeurs cible et la moyenne

Indicateurs de performance

Pour chaque cible

rr_<TargetName>	Nom de la cible
kc_<TargetName>	Il est à noter que rr_ indique une régression ou une classification et kc_ une segmentation.
Capacité prédictive (KI)	Indicateur de qualité qui correspond à la proportion d'information contenue dans la variable cible que les variables explicatives peuvent expliquer.
Reproductibilité (KR)	Indicateur de robustesse qui précise la capacité du modèle à obtenir les mêmes performances lorsqu'il est appliqué à un nouveau jeu de données ayant les mêmes caractéristiques que le jeu de données d'entraînement.

Le graphique suivant affiche la capacité prédictive et la reproductibilité :



Segmentation

Nombre de segments

Pour chaque variable cible

<nom de la cible>	Nom de la variable cible

Nombre de segments demandés	Nombre de segments demandés par l'utilisateur
Nombre de segments trouvés	Nombre de segments trouvés par le moteur

Indicateurs de performance supplémentaires

Pour les modèles de segmentation, vous pouvez évaluer votre modèle en contrôlant les indicateurs de performance par segment ou par jeu de données.

Voici quelques informations préliminaires :

- C_k est le *K*ème segment.
- x_i est un individu appartenant à C_k.
- w_i est le poids de x_i. Il s'agit de la valeur de la colonne "Poids" du jeu de données ou de la valeur 1 par défaut.
- μ_k est le centroïde de C_k.
- $W_k = \sum_{x_i \in C_k} w_i$ est le poids total de C_k.
- $W = \sum_k W_k$ est le poids total du jeu de données.
- $d(x_i, x_j)$ est la fonction de distance (euclidienne, de Manhattan, etc.) entre x_i et x_j.

Les indicateurs de performance disponibles par segment sont les suivants :

Indicateur	Définition	Formule
Somme des carrés résiduels (RSS)	Somme des distances quadratiques entre le centroïde et chaque observation individuelle.	$\sum_{x_i \in C_k} w_i \, d^2(x_i, \mu_k)$
Variance interne (aussi appelée inertie intra-classe)	Moyenne des distances quadratiques entre le centroïde et chaque observation individuelle.	$\frac{1}{W_k} \sum_{x_i \in C_k} w_i \, d^2(x_i, \mu_k)$
Silhouette simplifiée	Indicateur dérivé du rapport moyen des distances de tous les individus à leurs deux centroïdes les plus proches	$1 - \frac{1}{W_k} \sum_{x_i \in C_k} w_i \frac{d(x_i, \mu_k)}{d(x_i, \pi_i)}$, où π_i est le second centroïde le plus proche de x_i.

Les indicateurs de performance agrégés par jeu de données sont les suivants :

Indicateur	Définition	Formule
Somme de la somme des carrés résiduels (RSSSum)	Somme de la somme des carrés résiduels du segment	$\sum_k \sum_{x_i \in C_k} w_i \, d^2(x_i, \mu_k)$
Variance interne (ou inertie intra-classe)	Somme des inerties intra-classe du segment	$\sum_k \frac{1}{W_k} \sum_{x_i \in C_k} w_i \, d^2(x_i, \mu_k)$
Silhouette simplifiée	Moyenne pondérée des indicateurs de la silhouette simplifiée du segment	$1 - \frac{1}{W} \sum_k \sum_{x_i \in C_k} w_i \frac{d(x_i, \mu_k)}{d(x_i, \pi_i)}$, où π_i est le second centroïde le plus proche de x_i.

5.5.1.4 Série temporelle

Variables cible continues

Pour chaque variable cible continue :

<nom de la cible>	Nom de la variable cible
Min	Valeur minimale de la variable cible dans le jeu de données d'entraînement
Max	Valeur maximale de la variable cible dans le jeu de données d'entraînement
Moyenne	Moyenne des valeurs de la variable cible pour le jeu de données d'entraînement
Écart-type	Moyenne de la distance entre les valeurs cible et la moyenne

Composantes du modèle

Un modèle est une combinaison contenant au moins un des trois types d'éléments suivants :

- Une *tendance*
- Un ou plusieurs *cycles*
- Une *fluctuation*

Sélection du modèle et indicateur de performance

MAE (erreur absolue moyenne, aussi appelée EAM)	Formule : $$L1 = \frac{1}{W}\sum_{i=1}^{N} w_i\, u_i$$ Où *W* est la pondération totale de la population, *Wi* est la pondération de l'observation testée et *Ui* est l'erreur.	**Que se passe-t-il lors de l'entraînement ?** Automated Analytics crée plusieurs modèles "candidats". Ensuite, le meilleur modèle de prévision de série temporelle est conservé. La sélection du modèle est basée sur la MAE globale sur l'horizon. **Comment la MAE globale sur l'horizon est-elle calculée ?** Le modèle de prévision de série temporelle est appliqué aux observations passées trouvées dans la partition de validation. Pour chaque période, le modèle calcule le nombre de valeurs prévues demandées. Il s'agit de l'horizon des prévisions. Chacune de ces valeurs prévues est comparée à la valeur réelle correspondante. Ensuite, pour chaque horizon possible, une MAE par horizon peut être calculée : La MAE globale sur l'horizon correspond à la moyenne de toutes les valeurs de MAE par horizon.
MAPE globale sur l'horizon	Formule permettant de calculer la moyenne des valeurs absolues de l'erreur relative (MAPE) : $$MAPE = \frac{1}{W}\sum_{i=1}^{N} w_i \frac{\lvert y_i - \hat{y}_i \rvert}{\lvert y_i \rvert}$$ Où W est la pondération totale de la population, Wi est la pondération des observations testées, Yi est la valeur du signal et $\hat{y}_i$ est le prédicteur.	Cet indicateur est la moyenne des valeurs MAPE observées dans tout l'horizon d'entraînement. Une valeur de zéro indique un modèle parfait tandis qu'une valeur supérieure à 1 indique un modèle de mauvaise qualité. Une valeur de 0,09 signifie que le modèle prend en compte 91 % du signal ou, en d'autres termes, que l'erreur de prévision relative (résidus du modèle) est de 9 %. MAPE - Moyenne des valeurs absolues de l'erreur relative : moyenne de la somme des valeurs absolues des erreurs relatives. Elle mesure la précision des prévisions du modèle et renseigne sur la qualité de l'adéquation entre les prévisions et le signal réel.

5.5.1.5 Règles d'association

Paramètres des jeux de données

Colonne ID de transaction	Colonne du jeu de données de transactions utilisée pour la jointure avec le jeu de données de référence
Colonne ID de référence	Colonne du jeu de données de référence utilisée pour la jointure avec le jeu de données d'événements
Colonne des éléments	Colonne de jeu de données contenant les éléments

Paramètres de l'entraînement

Support minimal	Le support d'une règle est le nombre d'enregistrements qui vérifient cette règle. Dans le cas d'une règle de forme X=>Y, le support est le nombre d'enregistrements contenant à la fois le jeu d'éléments X et l'attribut Y. Le support minimum est le nombre minimum d'enregistrements nécessaires pour conserver une règle.
Confiance minimum	La confiance d'une règle est le pourcentage d'enregistrements qui vérifient le conséquent de la règle parmi ceux qui vérifient son antécédent. Dans le cas d'une règle de la forme X=>Y, la confiance de la règle est le nombre d'enregistrements contenant à la fois le jeu d'éléments X et l'attribut Y comparé au nombre d'enregistrements contenant uniquement le jeu d'éléments X.
Taille maximum des règles	La taille maximum d'une règle est le nombre total d'éléments qu'elle contient, y compris l'antécédent et le conséquent. On appelle ça la cardinalité d'une règle.

Résultats

Nombre de sessions traitées	Nombre de sessions analysées
Nombre de règles trouvées	Nombre de règles trouvées par les règles d'association
Nombre d'éléments trouvés	Nombre d'éléments trouvés
Jeux d'éléments fréquents	Nombre de jeux d'éléments fréquents, c'est-à-dire le nombre de jeux d'éléments dont le support est supérieur au support minimum.
Transactions traitées	Nombre de transactions analysées par les règles d'association

5.5.1.6 Journalisation des événements et analyseur de séquences

Agrégation de données

Données des événements	Nom de la table ou du fichier contenant les données des événements
Nombre total de références	Nombre d'enregistrements dans le jeu de données de référence
Nombre de références concordantes	Nombres de références trouvées à la fois dans le jeu de données de référence et dans celui des événements
Nombre de transactions traitées	Nombre de transactions analysées par le moteur
Nombre total de transactions	Nombre de transactions trouvées dans le jeu de données des événements

5.5.1.7 Social

Tous les réseaux

Nombre de graphes	Le nombre de graphes qui ont été créés
Nœuds dans le premier référentiel	Le nombre de nœuds qui ont été trouvés ou, dans le cas d'un graphe bipartite, le nombre de nœuds dans la première population
Nœuds dans le deuxième référentiel	Le nombre de nœuds dans la deuxième population (uniquement pour des graphes bipartites)

Détail des réseaux

Pour chaque graphe qui a été créé :

<Nom>	Nom du graphe
Type de liens	Le type de liens dans le graphe (dirigé ou non dirigé)
Nombre de nœuds	Le nombre de nœuds qui ont été trouvés
Nombre de liens	Le nombre de liens entre les nœuds
Densité	Rapport entre le nombre de liens existants et le nombre total de liens possibles

5.5.1.8 Recommandation

Détails

Pour chaque recommandeur, les informations suivantes sont fournies :

Exemple de code

```
<Name of the recommender>
```

Transactions

Nb. d'utilisateurs distincts	Nombre d'utilisateurs distincts identifiés dans le jeu de données d'entraînement
Nb. d'éléments distincts	Nombre d'utilisateurs distincts identifiés dans le jeu de données d'entraînement
Nb. de paires utilisateur/élément	Nombre de paires distinctes reliant un utilisateur et un élément
Meilleures ventes	Nombre d'éléments identifiés comme "meilleures ventes", c'est-à-dire qui apparaissent dans les transactions un nombre de fois supérieur au seuil défini dans la section "Identifier les meilleures ventes" lors du paramétrage du modèle.

5.5.2 Encodage textuel - Apprentissage

Une fois les paramètres définis, l'encodage textuel analyse et calcule les données. Une barre de progression vous permet de suivre le processus.

Lorsque le traitement est achevé, une table d'analyse s'affiche. Pour chaque variable textuelle, les racines conservées dans le dictionnaire sont listées avec leur nombre d'apparition et leur fréquence en pourcentage.

Vérifiez les racines et si la sélection des racines répond à vos attentes, cliquez sur le bouton *Suivant*. Sinon cliquez sur le bouton *Précédent* et modifiez les paramètres.

5.5.3 Vérification du modèle

À ce stade, l'application analyse les jeux de données et crée un certain nombre de variables ou de colonnes. En fonction des opérations que vous choisissez au cours de l'étape précédente, l'analyseur de séquences crée :

- quatre colonnes standard - *ksc_Start_Date*, *ksc_End_Date*, *ksc_TotalTime* et *ksc_Number_Events*,

- une colonne pour chaque état (dans le cas de l'option *Count*),
- une colonne pour chaque transition (si vous avez sélectionné **Compter les transitions**).
- Deux colonnes, *Premier état* et *Dernier état* (si vous avez sélectionné *Compter*, *Compter les transitions* ou *Premier et dernier*).
- Six colonnes, *LastStepNumber*, *Last_date-time*, *Last_duration*, *Session_Continue*, *LastState*, et *NextState* (si vous avez sélectionné *Séquences intermédiaires*).

Lorsque la vérification du modèle est terminée, vous pouvez vérifier le nombre de colonnes créées en regardant le détail du processus.

5.6 Utilisation du modèle

5.6.1 Classification/Régression

Cet écran affiche les différentes options proposées pour visualiser, appliquer, enregistrer ou exporter le modèle.

Affichage

- *Aperçu du modèle* : affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.
- *Courbes de performance* : selon le type de cible, le graphique des courbes de performances vous permet de :
  - Visualiser le profit réalisable par rapport à votre problématique en utilisant le modèle généré lorsque la cible est nominale.
  - Comparer les performances du modèle généré à celles d'un modèle aléatoire et d'un modèle parfait hypothétique quand la variable cible est nominale.
  - Comparer la valeur prédite à la valeur réelle dans le cas d'une variable cible continue.
- *Contributions par variables* : ces graphes montrent l'importance de chaque variable pour prédire la variable cible. Plus une variable est située à gauche dans le graphique et plus elle contient d'informations servant à la prédiction de la variable cible.
- *Importance des catégories* : ce panneau détaille l'importance de chaque catégorie pour les variables nominales ou ordinales et l'importance de chaque classe pour les variables continues.
- *Corrélations :* ce panneau affiche les variables corrélées ainsi que leur degré de corrélation.
- *Rapports statistiques détaillés (HTML) :* pour vous aider à analyser les résultats de votre modèle et vous permettre de partager ces résultats avec vos collègues, vos responsables ou vos clients, vous pouvez générer des rapports statistiques détaillés au format HTML. .
- *Tables d'analyse* : en plus des *rapports statistiques détaillés*, vous obtenez un ensemble de tables vous permettant d'analyser plus finement votre modèle.
- *Scorecard* : Influence et coefficients de chaque catégorie pour toutes les variables du modèle.

- *Paramètres du modèle* : toutes les informations nécessaires au codage des variables, la compression, le remplacement des variables manquantes, la fréquence des variables, les paramètres du modèle, etc., sont stockées dans un arbre des paramètres.

Exécution

- *Vérification des écarts* : l'utilisateur peut charger un nouveau jeu de données et vérifier s'il existe des écarts par rapport au jeu de données ayant été utilisé pour créer le modèle.
- *Appliquer le modèle à un nouveau jeu de données* : un nouveau jeu de données peut être scoré en utilisant le modèle créé. Les scores peuvent également être appliqués aux données utilisées pour créer le modèle.
- *Simuler le modèle* avec de nouvelles valeurs : en utilisant le modèle généré, l'utilisateur peut simuler le score et la probabilité pour un cas spécifique.
- *Sélectionner les variables à forte contribution :* un utilisateur peut sélectionner un sous-ensemble des variables les plus importantes pour prédire la variable cible.

Enregistrement/Export

- *Génération du code :* le codage des variables et les paramètres du modèle peuvent être stockés en tant que code qui peut être compilé et utilisé sur d'autres systèmes pour modéliser des données sans utiliser SAP Predictive Analytics.
- *Export KxShell :* génère un script KxShell qui peut être utilisé pour recréer le modèle.
- *Enregistrer le modèle en cours* : le modèle créé peut être enregistré et rechargé ultérieurement.

5.6.2 Mise en cluster

Ce panneau affiche les différentes options proposées pour visualiser, appliquer et enregistrer ou exporter le modèle de mise en cluster.

Affichage

- *Aperçu du modèle* : affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.
- *Courbes de performance* : selon le type de cible, le graphique des courbes de performances vous permet de :
  - Visualiser le profit réalisable par rapport à votre problématique en utilisant le modèle généré lorsque la cible est nominale.
  - Comparer les performances du modèle généré à celles d'un modèle aléatoire et d'un modèle parfait hypothétique quand la variable cible est nominale.

    - Comparer la valeur prédite à la valeur réelle dans le cas d'une variable cible continue.
- *Importance des catégories* : ce panneau détaille l'importance de chaque catégorie pour les variables nominales ou ordinales et l'importance de chaque classe pour les variables continues.
- *Graphiques des clusters* : affiche un diagramme à barres des clusters triés par fréquence, moyenne de la variable cible, ou moyenne relative.
- *Statistiques croisées :* compare chaque cluster à l'ensemble de la population pour les catégories de chaque variable.
- *Rapports statistiques détaillés (HTML) :* pour vous aider à analyser les résultats de votre modèle et vous permettre de partager ces résultats avec vos collègues, vos responsables ou vos clients, vous pouvez générer des rapports statistiques détaillés au format HTML.
- *Tables d'analyse* : en plus des *rapports statistiques détaillés*, vous obtenez un ensemble de tables vous permettant d'analyser plus finement votre modèle.
- *Paramètres du modèle* : toutes les informations nécessaire au codage des variables, la compression, le remplacement des valeurs manquantes, la fréquence des variables, les paramètres du modèle, etc. sont stockés dans l'arbre des paramètres.

Exécution

- *Vérification des écarts* : l'utilisateur peut charger un nouveau jeu de données et vérifier s'il existe des écarts par rapport au jeu de données ayant été utilisé pour créer le modèle.
- *Appliquer le modèle à un nouveau jeu de données* : à l'aide du modèle, il est possible d'attribuer des définitions de clusters aux enregistrements d'un nouveau jeu de données.
- *Simuler le modèle* avec de nouvelles valeurs : en utilisant le modèle généré, l'utilisateur peut simuler le score et la probabilité pour un cas spécifique.

Enregistrement/Export

- *Export KxShell :* génère un script *KxShell* qui peut être utilisé pour recréer le modèle.
- *Génération du code :* le codage des variables et les paramètres du modèle peuvent être stockés en tant que code qui peut être compilé et utilisé sur d'autres systèmes pour modéliser des données sans utiliser SAP Predictive Analytics.
- *Enregistrer le modèle en cours* : le modèle créé peut être enregistré et rechargé ultérieurement.

5.6.3 Série chronologique

Cet écran affiche les différentes options proposées pour visualiser, appliquer et enregistrer le modèle de série temporelle.

Affichage

- *Aperçu du modèle* : contient des informations sur le modèle généré telles que le nombre de variables, les informations sur le signal, les composantes du modèle, etc.
- *Afficher les prévisions* : vous permet de visualiser les prévisions modélisées, c'est-à-dire le signal et le signal prévu, la tendance et les intervalles de prédiction.
- *Afficher les composantes du signal* : contient les informations sur les composantes du modèle, c'est-à-dire les composantes du polynôme utilisé pour générer les prévisions (tendance, cycles, ...).
- *Régressions : Contributions par variables* : vous permet de visualiser les contributions des variables pour toutes les régressions utilisées dans le modèle.
- *Tables d'analyse* : ensembles de tableaux vous permettant une analyse détaillée de votre modèle.
- *Paramètres du modèle* : toutes les informations nécessaires au codage des variables, la compression, le remplacement des valeurs manquantes, la fréquence des variables, les paramètres du modèle, etc. sont stockées dans l'arbre des paramètres.
- *Appliquer le modèle* : vous permet d'enregistrer les prévisions générées dans un fichier.
- *Enregistrer le modèle en cours* : le modèle créé peut être enregistré et rechargé ultérieurement.

5.6.4 Règles d'association

Cet écran affiche les différentes options proposées pour visualiser les règles d'association, appliquer et enregistrer le modèle.

Affichage

- *Aperçu du modèle* : affiche des informations sur le modèle créé comme le nombre de règles d'association trouvées, un rappel des paramètres du modèle, etc.
- *Exploration des règles d'association* : permet d'effectuer des recherches parmi toutes les règles d'association trouvées en utilisant des options de filtrage.
- *Tables d'analyse* : ensembles de tableaux vous permettant une analyse détaillée de votre modèle.
- *Paramètres du modèle* : toutes les informations nécessaire au codage des variables, la compression, le remplacement des valeurs manquantes, la fréquence des variables, les paramètres du modèle, etc. sont stockés dans l'arbre des paramètres.

5.6.5 Options autonomes de journalisation des événements et d'analyseur de séquences

Cet écran affiche les différentes options proposées pour afficher, appliquer et enregistrer ou exporter le modèle.

Affichage

- *Aperçu du modèle* : affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.
- *Tables d'analyse* : outre les statistiques descriptives, SAP Predictive Analytics vous fournit un ensemble de tables présentant des statistiques plus détaillées de votre modèle.
- *Paramètres du modèle* : Toutes les informations nécessaires au codage des variables, la compression, le remplacement des variables manquantes, la fréquence des variables, les paramètres du modèle, etc., sont stockées dans un arbre des paramètres.

Exécution

- *Application du modèle à un nouveau jeu de données* : un nouveau jeu de données peut être scoré en utilisant le modèle créé. Les scores peuvent également être appliqués aux données utilisées pour créer le modèle.
- *Enregistrer le modèle en cours* : le modèle créé peut être enregistré et rechargé ultérieurement.

5.6.6 Options d'encodage textuel seul

Cet écran affiche les différentes options proposées pour afficher, appliquer et enregistrer ou exporter le modèle.

Affichage

- *Aperçu du modèle* : affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.
- *Tables d'analyse* : en plus des rapports statistiques détaillés, vous obtenez un ensemble de tables vous permettant d'analyser plus finement votre modèle.
- *Paramètres du modèle* : Toutes les informations nécessaires au codage des variables, la compression, le remplacement des variables manquantes, la fréquence des variables, les paramètres du modèle, etc., sont stockées dans un arbre des paramètres.

Exécution

- *Application du modèle à un nouveau jeu de données* : un nouveau jeu de données peut être scoré en utilisant le modèle créé. Les scores peuvent également être appliqués aux données utilisées pour créer le modèle.

- *Exporter le script KxShell* : génère un script KxShell qui peut être utilisé pour recréer le modèle.
- *Enregistrer le modèle en cours* : le modèle créé peut être enregistré et rechargé ultérieurement.

5.6.7 Mode Sans application ou Simulation

Cette section décrit les options disponibles pour les modèles de classification/régression et de segmentation/mise en cluster lors de l'utilisation du mode **Sans application ou Simulation**.

5.6.7.1 Classification/Régression

Cet écran affiche les différentes options proposées pour visualiser, appliquer, enregistrer ou exporter le modèle.

Affichage

- *Aperçu du modèle* : affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.
- *Courbes de performance* : selon le type de cible, le graphique des courbes de performances vous permet de :
  - Visualiser le profit réalisable par rapport à votre problématique en utilisant le modèle généré lorsque la cible est nominale.
  - Comparer les performances du modèle généré à celles d'un modèle aléatoire et d'un modèle parfait hypothétique quand la variable cible est nominale.
  - Comparer la valeur prédite à la valeur réelle dans le cas d'une variable cible continue.
- *Contributions par variables* : ces graphes montrent l'importance de chaque variable pour prédire la variable cible. Plus une variable est située à gauche dans le graphique et plus elle contient d'informations servant à la prédiction de la variable cible.
- *Importance des catégories* : ce panneau détaille l'importance de chaque catégorie pour les variables nominales ou ordinales et l'importance de chaque classe pour les variables continues.
- *Corrélations :* ce panneau affiche les variables corrélées ainsi que leur degré de corrélation.
- *Rapports statistiques détaillés (HTML) :* pour vous aider à analyser les résultats de votre modèle et vous permettre de partager ces résultats avec vos collègues, vos responsables ou vos clients, vous pouvez générer des rapports statistiques détaillés au format HTML. .
- *Tables d'analyse* : en plus des *rapports statistiques détaillés*, vous obtenez un ensemble de tables vous permettant d'analyser plus finement votre modèle.
- *Scorecard* : Influence et coefficients de chaque catégorie pour toutes les variables du modèle.
- *Paramètres du modèle* : toutes les informations nécessaires au codage des variables, la compression, le remplacement des variables manquantes, la fréquence des variables, les paramètres du modèle, etc., sont stockées dans un arbre des paramètres.

Exécution

- *Vérification des écarts* : l'utilisateur peut charger un nouveau jeu de données et vérifier s'il existe des écarts par rapport au jeu de données ayant été utilisé pour créer le modèle.
- *Appliquer le modèle à un nouveau jeu de données* : un nouveau jeu de données peut être scoré en utilisant le modèle créé. Les scores peuvent également être appliqués aux données utilisées pour créer le modèle.
- *Simuler le modèle* avec de nouvelles valeurs : en utilisant le modèle généré, l'utilisateur peut simuler le score et la probabilité pour un cas spécifique.
- *Sélectionner les variables à forte contribution :* un utilisateur peut sélectionner un sous-ensemble des variables les plus importantes pour prédire la variable cible.

Enregistrement/Export

- *Génération du code :* le codage des variables et les paramètres du modèle peuvent être stockés en tant que code qui peut être compilé et utilisé sur d'autres systèmes pour modéliser des données sans utiliser SAP Predictive Analytics.
- *Export KxShell :* génère un script KxShell qui peut être utilisé pour recréer le modèle.
- *Enregistrer le modèle en cours* : le modèle créé peut être enregistré et rechargé ultérieurement.

5.6.7.2 Segmentation/Mise en cluster

Ce panneau affiche les différentes options proposées pour visualiser, appliquer et enregistrer ou exporter le modèle de mise en cluster.

Affichage

- *Aperçu du modèle* : affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.
- *Courbes de performance* : selon le type de cible, le graphique des courbes de performances vous permet de :
  - Visualiser le profit réalisable par rapport à votre problématique en utilisant le modèle généré lorsque la cible est nominale.
  - Comparer les performances du modèle généré à celles d'un modèle aléatoire et d'un modèle parfait hypothétique quand la variable cible est nominale.
  - Comparer la valeur prédite à la valeur réelle dans le cas d'une variable cible continue.
- *Importance des catégories* : ce panneau détaille l'importance de chaque catégorie pour les variables nominales ou ordinales et l'importance de chaque classe pour les variables continues.
- *Graphiques des clusters* : affiche un diagramme à barres des clusters triés par fréquence, moyenne de la variable cible, ou moyenne relative.

- *Statistiques croisées :* compare chaque cluster à l'ensemble de la population pour les catégories de chaque variable.
- *Rapports statistiques détaillés (HTML) :* pour vous aider à analyser les résultats de votre modèle et vous permettre de partager ces résultats avec vos collègues, vos responsables ou vos clients, vous pouvez générer des rapports statistiques détaillés au format HTML.
- *Tables d'analyse* : outre les statistiques descriptives, SAP Predictive Analytics vous fournit un ensemble de tables présentant des statistiques plus détaillées de votre modèle.
- *Paramètres du modèle* : toutes les informations nécessaire au codage des variables, la compression, le remplacement des valeurs manquantes, la fréquence des variables, les paramètres du modèle, etc. sont stockés dans l'arbre des paramètres.

Exécution

- *Vérification des écarts* : l'utilisateur peut charger un nouveau jeu de données et vérifier s'il existe des écarts par rapport au jeu de données ayant été utilisé pour créer le modèle.

Enregistrement/Export

- *Export KxShell :* génère un script KxShell qui peut être utilisé pour recréer le modèle.
- *Génération du code :* le codage des variables et les paramètres du modèle peuvent être stockés en tant que code qui peut être compilé et utilisé sur d'autres systèmes pour modéliser des données sans utiliser SAP Predictive Analytics.
- *Enregistrer le modèle en cours* : le modèle créé peut être enregistré et rechargé ultérieurement.

5.6.8 Analyse de la vérification des écarts

Cet écran permet d'afficher les analyses détaillées de la vérification des écarts :

Le panneau...	affiche...
Probabilité d'écart	la probabilité d'écart de chaque variable par rapport au jeu de données initial.
Probabilité d'écart par rapport à la cible	la probabilité d'écart de chaque catégorie de chaque variable en relation avec chacune des cibles du modèle par rapport au jeu de données initial.
Probabilité d'écart par catégorie	la probabilité d'écart de chaque catégorie de chaque variable par rapport au jeu de données initial.
Probabilités d'écart par catégories groupées	la probabilité d'écart de chaque groupe de catégories pour chaque variable par rapport au jeu de données initial.

Une valeur de 0,95 ou plus est le signe d'une forte probabilité d'écart qui mérite d'être étudiée de plus près.

5.7 Affichage d'un modèle

5.7.1 Affichage de la classification/régression

Cette section décrit les options vous permettant de générer un rapport détaillé sur un modèle de classification ou de régression.

Informations associées

Rapports statistiques [page 171]
Courbes de performance [page 176]
Détails relatifs aux variables [page 180]
Contributions par variables [page 182]
Arbre de décision [page 183]
Matrice de confusion [page 185]
Scorecard [page 187]

5.7.1.1 Rapports statistiques

Cet écran vous permet de visualiser des tableaux d'analyse détaillés du modèle selon le type de modèle que vous avez généré :

- Statistiques descriptives (classification/régression, segmentation)
- Performances du modèle (classification/régression, segmentation)
- Détails des segments (segmentation uniquement)
- Analyse avancée (classification/régression, segmentation)
- Indicateurs de performance (série temporelle)
- Variables cycliques (série temporelle)

Statistiques descriptives

Variables	Pour chaque variable : • son nom (*Variable*) • son type de valeur (*Valeur*) • son type de stockage (*Stockage*) • • le nombre de valeurs manquantes dans le jeu de données d'entraînement (*Nombre de valeurs manquantes*) • son rôle dans la modélisation (*Rôle*)
Fréquence des catégories	Pour chaque variable : • Catégories • Proportion de chaque catégorie dans le jeu de données complet
Variables continues	Pour chaque variable continue dans chacun des jeux de données (entraînement, validation, test) : • Min • Max • Moyenne • Variance (mesure de la dispersion) • Écart-type (mesure de l'étendue de la dispersion des nombres autour de leur moyenne, c'est-à-dire la racine carrée de la variance)

Statistiques croisées avec la cible

Statistiques croisées avec cibles nominales	Statistiques croisées entre les variables d'entrée et les variables cible nominales. C'est-à-dire, pour chaque variable croisée avec chaque cible nominale : • les catégories de variable • la proportion de la catégorie non-cible de la cible sélectionnée dans la catégorie de variable • la proportion de la catégorie cible de la cible sélectionnée dans la catégorie de variable • la fréquence de la catégorie en cours dans le jeu de données

Statistiques croisées avec cibles continues	Statistiques croisées entre les variables d'entrée et les variables cible continues. C'est à dire, pour chaque variable croisée avec chaque variable cible continue : • les catégories de variable • la moyenne de la variable cible pour chaque catégorie de la variable en cours • la variance de la variable cible pour chaque catégorie de la variable en cours

Performances du modèle

Capacité prédictive et reproductibilité	Pour chaque variable cible dans chacun des jeux de données (entraînement, validation, test) : • Nom de la variable • Valeur de la capacité prédictive • Valeur de reproductibilité
Contribution maximale intelligente des variables	Pour chaque variable cible : • Variables d'entrée • Contribution de chaque variable d'entrée
Autres indicateurs de performance	Pour chaque variable cible dans chacun des jeux de données (entraînement, validation, test) : • erreur absolue moyenne (*L1*) • racine de l'erreur quadratique moyenne (*L2*) • erreur maximale (*LInf*) • erreur moyenne (*ErrorMean*) • écart-type de l'erreur (*ErrorStdDev*) • taux de classification (*ClassificationRate*)

Détails des segments

Statistiques des segments pour variables nominales	Statistiques croisés standard des segments avec les variables nominales
Statistiques des segments pour variables nominales (normalisées)	Statistiques croisées normalisées des segments avec les variables nominales
Statistiques des segments pour variables continues	Statistiques croisées des segments avec les variables continues

Indicateurs KL des segments	Pour chaque segment :
	• mesure de Kullback-Leibler entre la population du segment et la population totale.
	  • Le *KL* (Kullback-Leibler) mesure la différence entre deux distributions. Plus le KL est élevé, plus les distributions sont différentes.
	  • Pour un segment et une dimension (ou variable) donnés, le KL est calculé grâce à la formule suivante : $\sum_{i=0}^{N} p_i \log\left(\frac{p_i}{q_i}\right)$
	  • où :
	  • N est le nombre de catégories pour cette dimension ;
	  • p i (ou q i) est la proportion de la ième catégorie (de la dimension concernée) dans le segment (ou la population totale). La proportion d'une catégorie donnée représente son poids dans le poids total du jeu de données concerné (soit le segment, soit la population globale).
	  • Le KL est utilisé dans le segment de statistiques croisées pour trier les variables par ordre décroissant.
	• Mesure *Khi carré* pour chaque variable d'entrée entre la population du segment et la population totale.
	  • Le khi carré est un autre indicateur de la divergence entre la distribution du segment et la distribution de la population pour cette dimension. Sa valeur représente le niveau de confiance que l'on peut avoir dans l'affirmation selon laquelle les deux distributions sont très différentes.
	  • À la différence du KL, il ne fournit pas d'indication sur l'ampleur des différences entre ces distributions. Il permet toutefois d'être sûr que les deux distributions SONT différentes lorsque sa valeur est proche de 1. Dans ce cas, nous pouvons considérer que la dimension concernée est réellement discriminative pour ce segment. Au contraire, une valeur égale à 0 suggère que les deux distributions sont probablement identiques, mais que cela reste une probabilité.

Fréquence des segments	Pour chaque variable cible dans chaque segment : • proportion d'élément du segment dans le jeu de données entier • proportion dans le segment de la valeur attendue de la cible ou moyenne de la cible

• Expressions SQL des segments	Pour chaque variable cible : • expression SQL des segments, si elle est disponible

Analyse Avancée

Groupes de catégories	Pour chaque variable par rapport à chaque variable cible : • Groupes de catégories • Proportion de chaque catégorie de la cible pour chaque groupe de la variable • Proportion de chaque groupe dans le jeu de données complet
Identifiant des groupes	Pour chaque variable par rapport à chaque variable cible : • Catégories Index du groupe contenant la catégorie actuelle L'absence d'index signifie que la catégorie n'est pas groupée.
Probabilité d'écart	Pour chaque variable et chaque jeu de données : • Probabilité d'écart de la variable par rapport à sa valeur dans le jeu de données d'entraînement
Probabilité d'écart par rapport à la cible	Pour chaque variable dans chacun des jeux de données par rapport à chaque variable cible : • Probabilité d'écart de la distribution de la variable cible par rapport à sa valeur dans le jeu de données d'entraînement
Autres indicateurs de performance des variables	Pour chaque variable dans chacun des jeux de données par rapport à chaque variable cible : • le K-S (*Ks*) • l'aire sous la courbe ROC (*AUC*) • le coefficient de GINI (*GINI*)
Codage continu	Pour chaque catégorie de chaque variable par rapport à chacune des variables cible : • la valeur de la catégorie • la valeur codée correspondante
Taille des jeux de données	Nombre d'enregistrements dans chacun des jeux de données (entraînement, validation et test)

Indicateurs de performance (pour InfiniteInsight® Modeler/Séries temporelles)

Le panneau...	*affiche...*
Intervalle de prédiction de la prévision	Pour chaque jeu de données : • intervalles de prédiction de la prévision calculées via la *racine de l'erreur quadratique moyenne (L2)*
efficacité de la prévision	Pour chaque prévision et chaque jeu de données : • l'efficacité de la prévision calculée sur *Pearson au carré* (P2), qui décrit la qualité du lien entre le prédicteur et le signal. Une valeur de 1 représente un lien parfait et donc un prédicteur parfait.
Autres indicateurs de performance	Pour chaque prévision et chaque jeu de données : • le *coefficient de qualité (U2)* (rapport entre la variance des résidus du prédicteur et la variance du signal) décrit si le modèle correspond au signal. Lorsque sa valeur est égale à 1, l'ajustement est parfait. • l'*erreur moyenne*, • l'*écart-type de l'erreur*.

Variables cycliques

Cycliques et saisonnales :

Pour chaque variable périodique trouvée par le modèle, le graphique affiche un cycle de cette variable.

> **i** Remarque
>
> Le nom d'une variable périodique correspond à son cycle. Par exemple, une variable nommée monthOfYear (mois de l'année) correspond au cycle des mois dans l'année. Alors qu'une variable nommée 52 correspond à un cycle de 52 unités (l'unité dépendant du signal).

5.7.1.2 Courbes de performance

Selon le type de cible, le graphique des courbes de performances vous permet de :

- Visualiser le profit réalisable par rapport à votre problématique en utilisant le modèle généré lorsque la cible est nominale.
- Comparer les performances du modèle généré à celles d'un modèle aléatoire et d'un modèle parfait hypothétique quand la variable cible est nominale.
- Comparer la valeur prédite à la valeur réelle dans le cas d'une variable cible continue.

Sur le graphique, les courbes représentent :

- lorsque la cible est nominale, le profit réalisable (en ordonnée) en fonction du rapport des observations correctement sélectionnées comme cibles sur la totalité du jeu de données initial (en abscisse) ;
- lorsque la cible est continue, la valeur prédite ou score (en abscisse) par rapport à la valeur réelle ou cible (en ordonnée).

- Modèles avec cible nominale [page 177]
- Modèles avec cible continue [page 179]

5.7.1.2.1 Modèles avec cible nominale

Affichage d'un graphe

Les abscisses représentent le pourcentage de la population initiale et les ordonnées le pourcentage du profit attendu maximum (y compris toute hypothèse coût/profit définie dans l'écran de la stratégie de profit).

Vous pouvez afficher uniquement la courbe de validation (par défaut) ou les courbes des trois ensembles de données. Le bouton *Afficher tous les jeux de données* vous permet de sélectionner les courbes Entraînement/Validation/Test.

La liste déroulante *Modèles* permet de sélectionner la variable cible pour laquelle vous voulez afficher les courbes.

La barre de menu située sous le titre vous permet de copier les données du graphique vers le presse-papier, d'imprimer le graphique ou de l'enregistrer au format PNG.

- La *courbe verte* représente le profit maximum possible (obtenue en utilisant la variable cible comme modèle). Par exemple, si 25 % de votre population appartiennent à la catégorie cible de la variable cible, le meilleur modèle classera correctement 25 % de la catégorie cible avec 25 % de la population.
- La *courbe rouge* représente le profit minimum (obtenue avec un modèle aléatoire). En prélevant de manière aléatoire 50 % de la population, vous identifieriez 50 % de la catégorie cible de la variable cible.
- La *courbe bleue* représente le profit généré en appliquant le modèle sur le jeu de données de validation. Cette courbe présente le lift de la courbe aléatoire basée sur le modèle.

Courbes de performances standard

La liste déroulante *Type d'analyse* permet de sélectionner le type de courbe à afficher. *Détecté* (option par défaut) est la courbe de profit standard alors que l'option *Normalisé* affichera une courbe de profit avec une moyenne de zéro. *Lift* vous permet d'étudier les différences entre un modèle parfait et un modèle aléatoire et celles entre le modèle généré par SAP Predictive Analytics et un modèle aléatoire.

La courbe de profit *Personnalisé* vous permet de spécifier les pertes pour la catégorie non-cible (par exemple le prix d'un courrier pour une personne ne répondant pas) et un gain pour la catégorie cible (par exemple un achat suite au courrier).

Courbes de performances avancées

Un ensemble de graphiques avancés sont proposés en plus des graphes standards décrits ci-dessus :

- La courbe *ROC* est dérivée de la théorie de la détection du signal.
  Elle mesure la performance de discrimination d'un modèle en matière de compromis entre sensibilité et spécificité ou, en fait, entre détection correcte et erronée, lorsque le seuil de détection varie.
  La *sensibilité*, qui figure sur l'axe Y, est la proportion de signaux CORRECTEMENT identifiés (vrais positifs) détectés (sur tous les vrais positifs du jeu de données de validation).
  [1 - Spécificité], qui figure sur l'axe X, est la proportion d'affectations INCORRECTES à la classe de signaux (faux positifs) (sur tous les faux positifs du jeu de données de validation). (Par opposition, la spécificité est la proportion de NON-SIGNAUX CORRECTEMENT identifiés (vrais négatifs).
- La courbe *Lorenz "Bon"* représente graphiquement les pourcentages cumulés de signaux manqués (faux négatifs) pour les enregistrements correspondant aux x % plus bas scores du modèle.
  Les ordonnées mesurent *[1 - sensibilité]*, c'est-à-dire [1 - proportion de vrais positifs], soit l'équivalent de la proportion de signaux manqués ou d'opportunités perdues. Comme les données sont classées des enregistrements les moins susceptibles d'être des signaux à gauche à ceux les plus susceptibles d'être des signaux à droite, plus la courbe monte lentement, plus le modèle est sensible en ce qui concerne les signaux détectés. La courbe Wizard se redresse à partir de l'axe des abscisses en un point correspondant au pourcentage de non-signaux dans le jeu de données de validation.
- La courbe *Lorenz "Mauvais"* représente graphiquement les pourcentages cumulés de vrais négatifs (spécificité) correspondant aux x% plus bas scores du modèle. Ici, plus la courbe monte vite, plus la fréquence de mauvaise détection est faible.
- Les *courbes de densité* représentent la fonction de densité du score des variables dans l'ensemble d'événements (courbe Densité “Bon”) et dans l'ensemble de non-événements (courbe Densité “Mauvais”). Ces courbes peuvent être également vues comme les "dérivées" des courbes de Lorenz (la fonction de densité est par définition la dérivée de la fonction de densité cumulée).
  La fonction de densité estimée sur une classe ou intervalle est égale à :
  (nombre d'événements dans l'intervalle/nombre total d'événements)/longueur de l'intervalle
  - La courbe *Densité "Bon"* représente la distribution des scores du modèle pour les signaux.
  - La courbe *Densité "Mauvais"* représente la distribution des scores du modèle pour les enregistrements qui ne sont pas des signaux.
  - La courbe *Densité "Tous"* représente sur le même graphique la distribution des scores du modèle pour les signaux et celle pour les non-signaux, ce qui vous permet de comparer les deux distributions.
- Les courbes de *"Risque"* représentent le score de risque, la densité de population et le rapport bon/mauvais. Ce rapport est calculé de la manière suivante :
  (1 - probabilité du risque)/probabilité du risque
  - La courbe *Densité de la population* représente le nombre d'enregistrements dans chaque segment de score de risque (20 par défaut).
  - La courbe *Probabilité du risque* représente la probabilité du risque et le score du risque.
  - La courbe *Probabilités Bon/Mauvais* représente la distribution du score du risque pour le rapport bon/mauvais.
  - La courbe *Log(Probabilités Bon/Mauvais)* représente la distribution du score du risque pour le rapport bon/mauvais des ordonnées logarithmiques.
  - La courbe *Risque "Tous"* rassemble les trois courbes de risque sur un seul graphe.

> **i** Remarque
>
> L'axe des ordonnées de la courbe de probabilité se trouve sur la droite. L'axe des ordonnées des courbes de densité de la population et Bon/Mauvais se trouve à gauche.

Formules

Le profil normal peut être calculé à l'aide de la formule suivante pour des cibles binaires si f1 est la fréquence de la classe cible la moins fréquente TC1 (et f2 = 1 - f1 est la probabilité de la classe cible la plus fréquente TC2). Dans ce cas, nous "associons" la classe cible la moins fréquente à un profit égal à f2 et la classe la plus fréquente à un profit égal à -f1. Ces profits normalisés sont déterminés de la façon suivante : `profit(TC1) * proba(TC1) + profit(TC2) * proba(TC2) = 0`

Ensuite on peut calculer : `normalProfit(C) = Profit(TC2) * P(TC2|C) + Profit(TC1) * P(TC1|C)`

5.7.1.2.2 Modèles avec cible continue

Le graphique par défaut affiche les valeurs de la *cible réelle* en fonction des valeurs de la *cible prédite*. Deux courbes sont représentées : une pour le sous-ensemble de *validation* (courbe bleue) et l'autre pour le modèle parfait hypothétique (*Wizard*, courbe verte). La courbe de *validation* représente la valeur de la *cible réelle* en fonction de la valeur de la *cible prévue*. Par exemple, lorsque le modèle prédit 35, la moyenne de la valeur réelle est 37. La courbe de l'*assistant* correspond simplement à `X=Y`, ce qui signifie que chaque valeur prédite est égale à la valeur réelle. Ce graphe permet de voir facilement et rapidement les erreurs du modèle. Lorsque la courbe s'éloigne trop du modèle parfait, cela signifie que la valeur prédite est suspecte.

Le graphe est calculé comme suit :

- Les valeurs prédites sont réparties dans environ 20 segments ou classes. Chacun de ses segments représente environ 5 % de la population.
- Pour chacun de ces segments, des statistiques basiques sont calculées sur la valeur réelle, telles que la moyenne du segment *(SegmentMean)*, la moyenne associée à la cible *(TargetMean)* et la variance de la cible sur ce segment *(TargetVariance)*. Par exemple, pour une valeur prédite `[17; 19]`, si la moyenne est égale à 18,5, la moyenne réelle est égale à 20,5 et la variance de la valeur réelle est égale à 9. Dans ce cas on peut dire que, si la valeur prédite se situe entre 17 et 19, le modèle sous-estime légèrement la valeur réelle.

Pour chaque courbe, un point est défini comme la moyenne d'un segment (SegmentMean) en abscisse et la moyenne associée à la cible en ordonnée (TargetMean).

La zone bleue représente l'écart standard attendu du modèle actuel. Cette zone représente environ 70 % des valeurs de la cible attendues.

Il est à noter que cet intervalle de prédiction (c'est-à-dire la moitié de la largeur de la zone bleue) est égal à une fois l'écart-type de la cible observée pour un segment de valeurs prédites. En d'autres termes, dans le cas d'une distribution gaussienne, cela signifie qu'environ 70 % des points réels devraient se trouver dans la zone bleue (il s'agit d'un pourcentage théorique qui n'est pas forcément observé à chaque fois). Le

paramètre par défaut du type de courbe est *Prédit/Réel*. Les valeurs extrêmes pour les plages de prédiction sont `{TargetMean - (sqrt(TargetVariance)); TargetMean + (sqrt(TargetVariance))}`

> **i** Remarque
>
> Il est à noter que sqrt(TargetVariance) correspond à l'écart-type et TargetMean +/- écart-type correspond à l'intervalle de confiance.

5.7.1.3 Profit personnalisé

Cet écran vous permet de spécifier la valeur du profit selon les catégories de la variable cible.

La colonne *Valeur possible* liste toutes les valeurs de la variable cible.

Dans la colonne *Profit*, associez un profit à chaque valeur de la cible. Le profit peut être positif ou négatif.

En général, un profit positif est associé à la valeur attendue de la cible et un profit négatif aux autres valeurs de la cible.

Quand vous avez fixé une valeur de profit pour chacune des catégories cible, cliquez sur *OK* pour afficher le graphique correspondant.

5.7.1.4 Détails relatifs aux variables

Cet écran représente l'importance relative des différentes catégories d'une variable par rapport à la variable cible. Vous pouvez afficher la courbe de profit de la variable sélectionnée en cliquant sur le bouton *Afficher la courbe de profit* situé dans la barre d'outils sous le titre.

L'importance d'une catégorie dépend à la fois de sa différence par rapport à la moyenne de la catégorie cible et du nombre de cas représentés. Une grande importance peut venir de :

- un écart significatif entre le nombre d'observation contenant la catégorie cible de la variable cible dans cette catégorie par rapport à la moyenne ;
- ou un écart mineur allié à un grand nombre d'observations dans cette catégorie ;
- ou une combinaison des deux.

La hauteur des barres indique le profit pour cette catégorie. Les barres positives correspondent aux catégories dont le nombre d'observations contenant la catégorie cible de la variable cible est au-dessus de la moyenne et les barres négatives correspondent aux catégories qui contiennent moins que la moyenne d'observations contenant la catégorie cible.

Le menu déroulant *Variables* vous permet de sélectionner et d'afficher le graphique pour chaque variable du modèle.

La barre de menu située sous le titre vous permet de copier les données du graphique vers le presse-papier, d'imprimer le graphique ou de l'enregistrer au format PNG.

Les valeurs sont normalisées et leur somme est toujours égale à 0. Selon la stratégie de profit choisie, ou dans le cas d'une variable cible de type continu, il peut arriver que les importances soient toutes positives ou toutes négatives.

- Les axes [page 181]
- Formules [page 181]
- Constante de normalisation (NC) [page 181]
- Profit standard (NP) [page 182]

5.7.1.4.1 Les axes

L'*axe des abscisses* représente l'influence des catégories de variable sur la cible. L'importance des différents nombres figurant en abscisse est expliquée dans le tableau suivant.

Nombre sur l'axe des abscisses	Indique que la catégorie a...
Nombre positif	une influence positive sur la cible.
0	aucune influence sur la cible (le comportement est identique au comportement moyen de la population totale).
Nombre négatif	une influence négative sur la cible.

L'*axe des ordonnées* représente les *catégories de variable*. Les catégories ayant le même effet sur la variable cible sont regroupées. Elles apparaissent comme suit : [Category_a;Category_b;Category_c]. Les catégories ne contenant pas suffisamment de données pour fournir des informations solides sont regroupées dans la catégorie KxOther. Lorsqu'une variable est associée à un trop grand nombre de valeurs manquantes, ces valeurs sont regroupées dans la catégorie KxMissing. Les deux catégories sont créées automatiquement.

5.7.1.4.2 Formules

Voici les formules utilisées pour l'importance des catégories :

```
Category Importance = NP*BF/{NC}
```

où *NP* est le *Profit standard*, *BF* est la *Fréquence de la classe* et *NC* est la *Constante de normalisation*.

5.7.1.4.3 Constante de normalisation (NC)

Le calcul de la constante de normalisation diffère selon le type de variable cible. Les calculs pour les variables cibles binaire et continues sont détaillés ci-dessous.

- Pour les cibles binaires :
  `(Target Frequency) * (1 - Target Frequency])`

- Pour les cibles continues :
  pour des cibles continues non pathologiques (c'est-à-dire des cibles continues sans pic de distribution (Dirac)) il peut être estimé comme suit :
  `(Proportion above Median) * (1-Proportion above Median) = 0.5 * (1 - 0.5) = 0.25`

5.7.1.4.4 Profit standard (NP)

Le profit standard peut être calculé en utilisant la formule suivante pour les cibles binaires.

Si on considère que f1 est la fréquence de la catégorie la moins fréquente de la variable cible, *TC1*, et que f2 = 1 - f1 est la fréquence de la catégorie la plus fréquente de la variable cible, *TC2*. On peut "associer" la catégorie la moins fréquente de la variable cible, c'est-à-dire *TC1* à un profit *profit(TC1)* égal à f2 et la catégorie la plus fréquente de la variable cible, c'est-à-dire *TC2*, à un profit profit(TC2) égal à -f1. Ces profits normalisés *profit(TC1)* et *profit(TC2)* sont déterminés afin que :

```
profit(TC1) * proba(TC1) + profit(TC2) * proba(TC2) = 0
```

Remarque : C'est la raison pour laquelle on parle de profil "standard".

On peut alors calculer le profit standard d'une catégorie C comme suit :

```
normalProfit(C) = Profit(TC2) * P(TC2|C) + Profit(TC1) * P(TC1|C)
```

Où *P(TC1|C)* est la probabilité conditionnelle d'appartenir à la catégorie la moins fréquente de la variable cible si l'individu appartient à la catégorie C. Cette probabilité conditionnelle est estimée par la fréquence de *TC1* parmi les individus appartenant à la catégorie C.

Le profit standard pour les cibles continues est une généralisation du cas binaire. Il repose seulement sur les fréquences des catégories de la cible et les probabilités conditionnelles. Pour en savoir plus, voir la documentation SAP Predictive Analytics.

5.7.1.5 Contributions par variables

Cet écran affiche plusieurs graphes montrant l'utilisation des différentes variables dans le modèle. Plus une variable est située à gauche dans le graphique et plus elle contient d'informations servant à la prédiction de la variable cible.

Quatre types de diagrammes sont disponibles :

- *Contribution des variables* : L'importance relative de chaque variable du modèle généré
- *Poids des variables* : Les poids (dans le polynôme final) des variables normalisées
- *Contributions intelligentes des variables* : ce graphe prend en compte la contribution interne des variables.
- *Contributions max. intelligentes des variables* : inclut uniquement le maximum des variables similaires. Par exemple, seul le codage des groupes de la variable continue *âge* sera affiché. C'est le diagramme affiché par défaut.

Les variables sont affichées sur l'*axe des abscisses*. Selon le type de diagramme, trois types de variables peuvent apparaître :

- la variable d'origine
- la variable codée par Automated Analytics Modeler (notée c_variable),
- la différence d'information entre deux variables corrélées (notée [variable_a]-[variable_b]).
  Lorsque deux variables sont corrélées, seule celle ayant la plus forte contribution est affichée, cependant il est possible qu'une partie de l'information diffère entre les deux variables. Dans ce cas, seule la différence est conservée.

L'*axe des ordonnées* représente la contribution des variables en fonction du diagramme sélectionné.

Pour modifier le type de diagramme, utilisez la liste déroulante *Type de diagramme*.

La barre de menu située sous le titre vous permet de copier les données du graphique vers le presse-papier, d'imprimer le graphique ou de l'enregistrer au format PNG.

Formules

Contribution intelligente des variables : l'équation calculée par Automated Analytics Modeler - Régression/ Classification peut s'écrire sous la forme $y = \sum_i (Ai * Xi)$, où `Ai` est le coefficient et `Xi` représente les valeurs d'entrée codées (`i` est le numéro des colonnes).

Poids des variables : le poids est simplement le coefficient associé à la variable normalisée. L'équation correspondant est la suivante : $y = \sum_i (Ai * Xi) = \sum_i (Wi * XNi)$ où `XN` est la version normalisée de `X`

5.7.1.6 Arbre de décision

Le panneau *Arbre de décision* vous permet d'afficher les résultats générés par le moteur de classification/ régression sous la forme d'un arbre de décision basé sur les cinq variables ayant la contribution la plus élevée.

1. Dans la liste déroulante *Cible*, sélectionnez la variable cible pour laquelle vous souhaitez afficher l'arbre de décision.

L'arbre de décision est affiché dans la partie supérieure du panneau. Chaque nœud de l'arbre indique :

- Le nom de la variable développée, par exemple Marital-status.
- Les catégories ayant servi à filtrer la population du nœud, par exemple {Married-AF-spouse;Never-married}.
- La *Population* totale du nœud.
- Le pourcentage de *Cible positive* (pour une cible nominale) ou la *Moyenne de la cible* (pour une cible continue).

Lorsque vous passez le curseur sur un nœud, plusieurs options sont disponibles :

	Sélectionner la variable à utiliser pour développer le niveau suivant de l'arbre de décision.
	Développer automatiquement le niveau suivant à l'aide de la variable ayant la plus forte contribution pas encore utilisée dans l'arbre de décision actuel.
	Replier la section de l'arbre affichée sous le nœud actuel

Lorsque vous sélectionnez un nœud, ses informations sont affichées dans l'onglet *Détails du nœud* situé dans la partie inférieure du panneau. Cet onglet vous fournit les informations suivantes :

- la cible pour laquelle l'arbre de décision est déployé.
- pour chaque jeu de données existant dans el modèle :

Population	nombre d'enregistrements existants pour le nœud
Pour des cibles continues	
Moyenne de la cible	moyenne de la cible pour le nœud
Pour des cibles nominales	
Cible positive	nombre d'enregistrements pour lesquels la cible est positive
Pourcentage de cible positive	pourcentage de la population du nœud pour laquelle la cible est positive
Cible négative	nombre d'enregistrements pour lesquels la cible est négative
Pourcentage de cible négative	pourcentage de la population du nœud pour laquelle la cible est négative
Variance	variance du nœud
Population pondérée	nombre d'enregistrements lorsque une variable de poids est utilisée

La courbe de profit pour l'arbre de décision est affichée dans l'onglet *Courbe de profit* (partie inférieure gauche du panneau). La courbe évolue en fonction des modifications faites sur l'arbre de décision.

Paramètres d'affichage

Le bouton *Paramètres d'affichage* vous permet de personnaliser certains paramètres d'affichage de l'arbre de décision.

- *Orientation* : ce paramètre vous permet de définir l'orientation de l'arbre, c'est-à-dire horizontale ou verticale.
- *Type d'affichage* : ce paramètre vous permet de choisir entre un affichage standard (*Compact*) et un affichage propre à SAP Predictive Analytics (*Complet*). L'option *Compact* occupe moins de place à l'écran mais elle est moins lisible que l'option *Complet*.

Une fois vos paramètres d'affichage définis, cliquez sur *Fermer*.

5.7.1.7 Matrice de confusion

Le panneau *Matrice de confusion* vous permet de comparer les valeurs cible prédites par le modèle avec les valeurs réelles et de définir le score au-dessus duquel les observations seront considérées comme positives : la valeur cible correspond à celle souhaitée. Ce panneau vous permet également de faire des simulations de profit selon le score choisi comme seuil ou d'adapter automatiquement le seuil pour obtenir un profit maximal.

Définitions

- On appelle **observation positive** toute observation appartenant à la *population cible*.
- On appelle **observation négative** toute observation n'appartenant pas à la *population cible*.

Paramètres

- Cible

Sélectionnez la variable cible pour laquelle vous voulez afficher la matrice de confusion en cliquant sur l'onglet correspondant. Seules les cibles nominales sont disponibles.

Matrice de confusion

Il existe trois façons de définir le seuil à l'aide du curseur affiché :

- en sélectionnant le *pourcentage de population visé* si la population est triée par ordre décroissant de score (*% de la population*)
- en sélectionnant le *pourcentage d'observations positives* que vous souhaitez détecter (*% de cible détectée*)
- en sélectionnant directement le *score* à utiliser comme seuil (*Score*). Toute observation dont le score est supérieur au seuil est considérée comme positive et toute observation dont le score est inférieur au seuil est considérée comme négative.

L'échelle est graduée du plus petit score (à gauche), au plus grand (à droite). Les valeurs correspondant à chaque option sont affichées dans des champs situés sous l'échelle.

Lorsque vous déplacez le curseur sur l'échelle, la matrice de confusion est modifiée en conséquence. Le tableau ci-dessous indique comment lire la matrice de confusion.

	`Predicted[Target Category]`	`Predicted[Non-target Category]`
	Observations positives prédites	Observations négatives prédites

Réel[Catégorie cible] Observations positives réelles	Nombre d'observations positives correctement prévues	Nombre d'observations réellement positives mais prédites négatives
Réel[Catégorie non-cible] Observations négatives réelles	Nombre d'observations réellement négatives mais prédites positives	Nombre d'observations négatives correctement prévues

Par défaut, la *Population totale* est égale au nombre d'enregistrements dans le jeu de données de validation. Vous pouvez modifier ce nombre pour visualiser la matrice sur la population sur laquelle vous voulez appliquer votre modèle.

- Les Métriques
- Le *Taux de classification* correspond à la proportion de données correctement classée par le modèle lors de son application sur le jeu de données d'entraînement.
- La *Sensibilité* d'un test mesure sa capacité à donner un résultat positif lorsqu'une hypothèse est vérifiée.
- La *Spécificité* d'un test mesure sa capacité à donner un résultat négatif lorsque l'hypothèse n'est pas vérifiée.
- La *Précision* est le pourcentage avec lequel des mesures répétées dans des conditions inchangées donnent des résultats identiques.
- Le *Score* indique à quel point la fonction de vraisemblance dépend de son paramètre.
  La fonction de vraisemblance est une fonction de probabilités conditionnelles qui décrit les valeurs d'une loi statistique en fonction de paramètres supposés connus.

Profit

Cette section vous permet de visualiser votre profit par rapport au score sélectionné, ou bien de sélectionner automatiquement le score d'après les paramètres de profit.

Pour chaque catégorie d'observations, saisissez un profit ou un coût par observation. Le profit total est automatiquement affiché dans le champ Profit situé à droite.

Pour sélectionner automatiquement le seuil de score qui vous permettra de faire le maximum de profit, remplissez la matrice de coût puis cliquez sur le bouton *Maximiser le profit*.

Exemple

Dans le tableau profit-coût ci-dessous, chaque observation positive correctement identifiée rapportera 15 € mais chaque observation négative identifiée comme positive coûtera 8 €.

Catégorie	Prévu[1]	Prévu[0]
Réel[1]	15	0
Réel[0]	-8	0

5.7.1.8 Scorecard

Cet écran vous indique les coefficients associés à chaque catégorie pour toutes les variables d'un *modèle de classification*.

- Pour obtenir un score
  - Additionnez tous les coefficients correspondant à la valeur de chaque variable sélectionnée.

5.7.2 Affichage de la segmentation/mise en cluster

Cette section décrit les options vous permettant de générer un rapport détaillé sur un modèle de segmentation ou de mise en cluster.

Informations associées

Profils des clusters [page 187]
Description détaillée des clusters [page 188]
Synthèse des clusters [page 188]

5.7.2.1 Profils des clusters

Cet écran affiche pour chaque variable les profils des clusters et de la population. La première moitié du panneau affiche des informations sur les clusters :

- *Nom du cluster* : nom du cluster.
- *Fréquences* : pourcentage de la population rassemblée dans le cluster.
- *% de "valeur"* : en cas de classification (cible binaire), il s'agit du pourcentage de la valeur dans le cluster, la "valeur" étant la catégorie clé de la cible binaire.

Dans la partie inférieure de l'écran, un graphique présente les statistiques croisées correspondant au cluster et à la variable. Des comparaisons peuvent être effectuées pour chaque catégorie d'une variable entre le cluster et l'ensemble de la population. En abscisses on trouve les catégories de la variable et en ordonnées, la population affichée en pourcentage. Par conséquent, pour chaque catégorie d'une variable, la proportion de cas présents dans un cluster est comparée à celle de l'ensemble de la population. Vous pouvez utiliser ces graphiques comme suit :

- pour un cluster donné, examinez chaque variable pour détecter les profils de variable qui diffèrent le plus de l'ensemble de la population.
- pour une variable donnée, examinez chaque cluster pour détecter quels profils de cluster diffèrent le plus de l'ensemble de la population.

Vous pouvez cocher la case *Fixer la Variable* pour afficher le graphe d'une variable donnée pour les différents clusters sélectionnés.

Lorsque vous sélectionnez un cluster dans le tableau situé dans la partie haute du panneau, celui-ci s'affiche dans la partie basse et peut être comparé à l'ensemble de la population.

Utilisez la liste déroulante *Variables* pour changer la variable à afficher dans la moitié inférieure de l'écran.

La barre de menu située sous le titre vous permet de copier les données du graphique vers le presse-papier, d'imprimer le graphique ou de l'enregistrer au format PNG.

5.7.2.2 Description détaillée des clusters

Cet écran affiche une description pour chaque cluster trouvé.

La partie supérieure de l'écran donne des statistiques de base pour chaque cluster :

- *Nom du cluster* : nom du cluster.
- *Fréquences* : pourcentage de la population rassemblée dans le cluster.
- % de "*valeur*" : pour les cas de classification (variable cible binaire), pourcentage de *valeur* dans le cluster, où *valeur* est la classe la moins fréquente de la variable cible.
- Moy. Var. cible : dans les cas de régression (variable cible continue), valeur moyenne de la variable cible pour les données affectées à ce cluster.

La partie inférieure de l'écran affiche les intervalles de valeurs rencontrées pour les données affectées à ce cluster ; les intervalles non- discriminants sont éliminés.

> **i** Remarque
>
> Dans la version actuelle, les intervalles ne peuvent pas être utilisés comme des instructions SQL pour sélectionner les données correspondant aux clusters (pour plus d'informations, voir la documentation liées à la fonctionnalité).

5.7.2.3 Synthèse des clusters

Cet écran affiche plusieurs types de graphes montrant des statistiques sur les clusters/segments :

- *Diagrammes à bulles*
  Les diagrammes à bulles affichent les clusters en représentant la relation entre trois variables. Un diagramme à bulles peut donc fournir 3 informations sur chaque cluster.
  L'axe des abscisses, l'axe des ordonnées et la taille de la bulle représentent chacun une variable.
  De plus, les diagrammes à bulles permettent une représentation graphique de la segmentation, vous permettant de visualiser les clusters.
  Utilisez les boutons placés sous la barre d'outils pour sélectionner les variables à utiliser dans le diagramme à bulles.

  > **i** Remarque
  >
  > Si vous avez choisi de générer l'expression SQL pour les clusters lors de la configuration des paramètres du modèle, un cluster supplémentaire s'affiche dans la synthèse des clusters. Ce cluster contient les enregistrements non affectés, c'est-à-dire les enregistrements qui ne peuvent pas être

décrits par les expressions SQL et qui sont laissées hors des clusters. Pour en savoir plus, voir les informations associées ci-dessous.

- *Diagrammes à barres (Graphiques des clusters)*
  Les diagrammes à barres affichent les trois *graphiques des clusters* :
  - Le graphique *Fréquences* présente le nombre d'observations contenues dans chaque cluster par rapport au nombre total d'observations contenues dans le jeu de données.
  - Le graphique *Moyenne de la cible* présente la proportion d'observations appartenant à la catégorie cible de la variable cible dans chaque cluster.
  - Comme le graphique *Moyenne de la cible*, le graphique *Moyenne relative de la cible* présente pour chaque cluster la proportion d'observations appartenant à la catégorie cible de la variable cible. La seule différence entre deux graphiques est l'échelle utilisée sur l'*axe des ordonnées*. Dans le graphique *Moyenne relative de la cible*, la valeur 0 de l'axe des ordonnées correspond au pourcentage réel d'observations appartenant à la catégorie cible de la variable cible par rapport au jeu de données entier.
    Le menu déroulant *Graphiques* vous permet de modifier le type de graphique.

La barre d'outils située sous la barre de titre vous permet de :

- passer d'un type de diagramme à l'autre,
- copier les coordonnées dans le presse-papier,
- imprimer les graphes,
- exporter sous format Excel,
- enregistrer les graphes au format PNG.

Informations associées

Statistiques croisées [page 126]

5.7.3 Exploration des règles d'association

Le panneau *Exploration des règles d'association* vous permet d'effectuer des recherches parmi toutes les règles trouvées par Automated Analytics en utilisant des options de filtrage.

Vous disposez de trois types de filtres :

- Par *Antécédent* : sélectionnez les antécédents que vous souhaitez voir apparaître dans les règles affichées
- Par *Conséquent :* sélectionnez les conséquents que vous souhaitez voir apparaître dans les règles affichées
- Par *filtres numériques* :
  - *Support de règle* (en nombre de sessions ou en pourcentage)
  - *Support de l'antécédent* (en nombre de sessions ou en pourcentage)
  - *Support du conséquent* (en nombre de sessions ou en pourcentage)
  - *KI* (capacité prédictive)
  - *Confiance*

- *Lift*

Vous pouvez soit utiliser un seul filtre, soit en combiner plusieurs pour affiner votre recherche.

L'association de plusieurs filtres vous permet de faire des sélections complexes de règles. Vous pouvez appliquer plusieurs sélections à des sous-ensembles différents, puis sélectionner des unions de ces sous-ensembles. Attention, ce genre d'opération nécessite que toutes les règles soient désélectionnées au préalable.

Vous pouvez aussi afficher les règles sous forme de graphe de façon à mieux visualiser les relations entre les jeux d'éléments.

Sélection des filtres

- Pour filtrer par antécédent et/ou conséquent
  - Sélectionnez les filtres que vous souhaitez appliquer :
    - Par *antécédent* : sélectionnez au moins un antécédent dans la liste.
    - Par *conséquent* : sélectionnez au moins un conséquent dans la liste. Si vous filtrez à la fois par antécédent et conséquent, veillez à ne pas sélectionner le même élément dans les deux listes.
- Pour utiliser un filtre numérique
  1. Sélectionnez un indicateur dans la liste de gauche (confiance, support de règle, taille, Lift, capacité prédictive (KI), support de l'antécédent, ...)
  2. Sélectionnez un opérateur de comparaison dans la liste du milieu.
  3. Saisissez une valeur de comparaison dans le champ de texte à droite.
  4. Cliquez sur le bouton > pour ajouter le filtre à la liste des filtres située complètement à droite
- Pour lancer la recherche
  - Une fois les filtres sélectionnés, cliquez sur le bouton *Rechercher*. La section *Règles d'association trouvées* affiche le nombre de règles trouvées et la liste de ces règles.

> **i** Remarque
>
> Si vous cliquez sur *Rechercher* sans avoir préalablement sélectionné de filtre, toutes les règles seront affichées. Si le nombre de règles est très élevé, leur affichage peut prendre du temps.

- Pour réinitialiser le moteur de recherche
  - Cliquez sur le bouton *Réinitialiser*.

Affichage des résultats

- Pour afficher les résultats de la recherche

Les résultats de la recherche sont affichés dans la partie inférieure de l'écran, dans la section *Règles d'association trouvées*.

1. Cochez les champs que vous souhaitez voir apparaître dans la liste de résultats. Les champs sélectionnés apparaissent automatiquement dans la liste.
2. Pour trier les résultats par colonne, cliquez sur le nom de la colonne choisie.

Activation ou désactivation des règles d'association

Toutes les règles d'association n'ont pas la même importance : certaines sont pertinentes tandis que d'autres n'apportent aucune information. C'est pourquoi Automated Analytics vous permet de désactiver certaines des règles d'association avant de les appliquer.

- Pour désactiver des règles d'association
  1. Créez un filtre.
  2. Cliquez sur le bouton *Rechercher* pour afficher un ensemble de règles d'association dans la section *Règles d'association trouvées*.
  3. Cliquez avec le bouton droit de la souris sur la liste des règles d'association. Un menu contextuel s'affiche.
     La première partie du menu, au-dessus du séparateur, vous permet de changer l'état d'activation des règles d'association :

L'option...	vous permet...
Activer les règles sélectionnées	d'activer les règles sélectionnées dans la liste lors de la prochaine application du modèle.
Désactiver les règles sélectionnées	de désactiver les règles sélectionnées dans la liste lors de la prochaine application du modèle.
Activer tout	d'activer toutes les règles d'association affichées dans la liste
Désactiver tout	de désactiver toutes les règles d'association affichées dans la liste

     La deuxième partie du menu, située sous le séparateur, vous permet de contrôler quelles règles d'association doivent être affichées dans la liste :
     - *Afficher toutes les règles* est l'affichage par défaut. Toutes les règles sont affichées, qu'elles soient actives ou non.
     - *Afficher seulement les règles actives*
     - *Afficher seulement les règles inactives*

     > **i** Remarque
     >
     > Les règles désactivées ne seront pas toutes utilisées lors de la prochaine application du modèle.

Afficher les règles sous forme de graphe

L'onglet *Graphe*, qui vous permet d'afficher les règles d'association sous forme d'un graphe, est composé de deux parties :

- La *partie gauche* liste *l'ensemble des éléments* utilisés dans au moins une règle, soit comme antécédent, soit comme conséquent,

- Le *graphe* est affiché dans la *partie droite* du panneau.
  Le tableau ci-dessous reprend les éléments utilisés dans le graphe :

L'élément...	Représente...
cercle bleu	un jeu d'éléments
taille des cercles	le support du jeu d'éléments représenté par le cercle
flèche	une règle
épaisseur des flèches	Par défaut, l'épaisseur des flèches représente la capacité prédictive (KI) de la règle, mais vous pouvez personnaliser ce paramètre (voir Paramétrage de l'affichage du graphe).
couleur des flèches	Par défaut, la couleur des flèches ne porte pas d'information, mais vous pouvez personnaliser ce paramètre (voir Paramétrage de l'affichage du graphe).

- Pour afficher les règles d'association sous la forme d'un graphe
  - Vous pouvez soit filtrer les règles trouvées par Automated Analytics et afficher le graphe correspondant :
    dans l'onglet *.. règle(s) trouvée(s)*, cliquez sur le bouton *Afficher les règles filtrées dans un graphe*.
  - soit choisir les antécédents et/ou les conséquents que vous souhaitez voir apparaître dans le graphe.
- Pour ajouter un élément au graphe
  1. Glissez-déposez l'élément sélectionné vers le graphe. Un menu contextuel s'affiche.
  2. Choisissez si vous voulez afficher les règles pour lesquelles l'élément fait partie de l'*antécédent* de la règle, du *conséquent* ou des deux :
     - si vous sélectionnez *Antécédent*, seules les règles dont l'antécédent contient cet élément seront affichées,
     - si vous sélectionnez *Conséquent*, seules les règles dont le conséquent contient cet élément seront affichées,
  - si vous sélectionnez *Antécédent ou conséquent*, seules les règles dont l'antécédent ou le conséquent contient cet élément seront affichées.
- Pour afficher les détails d'une règle
  - Passez la souris sur la flèche correspondant à la règle pour laquelle vous souhaitez afficher les détails (capacité prédictive, Lift, support, ...). Une fenêtre contenant les informations s'affiche.

Paramétrage de l'affichage du graphe

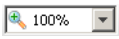
Vous pouvez paramétrer l'épaisseur et/ou la couleur des flèches représentant les règles.

- Pour paramétrer l'affichage des règles
  - Cliquez sur le bouton (*Personnaliser l'affichage des règles*). Une fenêtre s'ouvre vous permettant de choisir les paramètres du graphe.
- Pour personnaliser l'épaisseur des flèches
  - Dans la liste *Épaisseur des flèches*, sélectionnez la valeur que vous souhaitez utiliser.
- Pour personnaliser la couleur des flèches
  - Dans la liste *Couleur des flèches*, sélectionnez la valeur que vous souhaitez utiliser. La couleur varie de vert à rouge, vert représentant la plus petite valeur et rouge la plus grande.
- Pour afficher la valeur utilisée comme épaisseur pour les règles

- Cochez l'option *Afficher la valeur utilisée pour l'épaisseur comme label*. La valeur correspondante est affichée pour chaque flèche.

Options

La barre d'outils située au-dessus du graphe vous permet de :

	l'enregistrer sous forme d'image PNG
	paramétrer l'affichage du graphe
	ajuster le graphe à la taille du panneau
	zoomer ou dézoomer
	Vider le graphe

5.7.4 Affichage des séries temporelles

Cette section décrit les options vous permettant de générer un rapport détaillé sur un modèle de séries temporelles.

Informations associées

Contributions par variables [page 193]
Affichage des fonctionnalités du signal [page 194]
Affichage des prévisions [page 196]

5.7.4.1 Contributions par variables

Le graphique *Contributions par variables* vous permet d'étudier l'importance relative de chacune des variables dans les régressions utilisées dans le modèle. Les variables et la variable cible utilisées dans la régression dépendent de la fonctionnalité à modéliser. Deux fonctionnalités peuvent être modélisées en utilisant une régression : la tendance et la fluctuation.

Le tableau suivant détaille les variables générées et les variables cible qui peuvent être utilisées dans les régressions :

Composante Modélisée	Variable cible possible	Variables utilisables par la régression
Tendance	• signal	fonctions de date, c'est à dire Time, square(Time), sqrt(Time) (où Time est une variable date)
Fluctuation	• signal • signal - tendance • signal - tendance - cycles	variables lag sur la cible

Les quatre types de graphiques suivants permettent de visualiser les contributions des variables :

- Contributions des variables
- Poids des variables
- Contributions intelligentes des variables
- Contribution maximale intelligente des variables

Le graphique...	Présente...
Contributions des variables	L'*importance relative* de chaque variable du modèle généré
Poids des variables	Les *poids* (dans le polynôme final) des variables normalisées
Contributions intelligentes des variables	Les contributions *internes* des variables.
Contribution maximale intelligente des variables	Les contributions internes des variables comprenant uniquement le *maximum de variables similaires*. Par exemple, seul le codage des groupes de la variable continue âge sera affiché.

5.7.4.2 Affichage des fonctionnalités du signal

Un modèle est une combinaison contenant au moins un des cinq types d'éléments suivants :

- Une *tendance*
- Un ou plusieurs *cycles*
- Une *fluctuation*
- *Des résidus*
- Des *valeurs anormales*.

- La tendance [page 195]
- Les cycles [page 195]
- La fluctuation [page 195]
- Les résidus [page 196]
- Les valeurs anormales [page 196]

5.7.4.2.1 La tendance

La tendance est l'orientation générale du signal. Il existe quatre types de tendances :

Type de tendance	Apparaît comme...
Polynomial	une courbe correspondant au polynôme détecté. Les fonctions du polynôme disponibles sont Time, square(Time), sqrt(Time), où Time représente la valeur du paramètre *DateColumnName*.
Linéaire	une ligne droite C'est un cas particulier de tendance *polynomiale*.
L1	un déplacement du signal d'un cran vers l'avant. Il s'agit de la prévision de base dans laquelle l'observation prévue est identique à la dernière observation du signal.
L2	un déplacement du signal de deux crans vers l'avant.

5.7.4.2.2 Les cycles

Les cycles sont des éléments périodiques qui doivent apparaître au moins deux fois dans le jeu de données d'*entraînement*. Les deux types de cycles existants sont détaillés dans le tableau ci-dessous :

Type de cycle	Description
Périodique	C'est un cycle qui ne dépend pas de la date. Un périodique est défini par le nombre d'unité de temps qu'il recouvre.
Saisonnier	C'est un cycle qui dépend de la date ou de l'heure. Par exemple, le jour dans le mois, l'heure de la journée, etc. **i Remarque** KTS peut détecter automatiquement les variables saisonnières suivantes : dayOfWeek (jour de la semaine), dayOfMonth (jour dans le mois), dayOfYear (jour dans l'année), weekOfYear (semaine de l'année), monthOfYear (mois de l'année), hourOfDay (heure de la journée).

5.7.4.2.3 La fluctuation

La fluctuation est ce qui reste quand la tendance et les cycles ont été retirés du signal. Elle est modélisée par une auto-régression utilisant une fenêtre de données antérieures pour modéliser le résidu actuel. Le nombre

d'observations dans la fenêtre est déterminé par KTS en fonction du nombre total d'observations contenues dans le jeu de données d'entraînement.

5.7.4.2.4 Les résidus

Les résidus sont ce qui reste quand la tendance, les cycles et la fluctuation ont été retirés du signal. Ils se composent de variations de signaux qui ne peuvent pas être modélisées.

5.7.4.2.5 Les valeurs anormales

Les valeurs anormales sont des points où la courbe prédictive est très éloignée de la courbe réelle. Elles sont représentées par un carré rouge sur le graphe.

Une valeur de signal réelle est qualifiée d'anormale lorsque l'erreur de prévision correspondante est considérée comme anormale par rapport à la moyenne observée pour le jeu de données d'estimation. L'indicateur d'erreur de prévision est la différence absolue entre les valeurs réelle et prédite. Elle est également appelée le résidu. Le seuil anormal du résidu est défini comme étant égal à trois fois l'écart standard des valeurs résiduelles pour un jeu de données d'estimation (ou de validation).

Une valeur Yi est considérée comme anormale si |Résidu(Yi) – Moyenne(Résidu(Y), Estimation)| > 3 *ÉcartType(Résidu(Y), Estimation).

Où :

- Résidu (Yi) = Yi – Prédiction(Yi)
- Yi = valeur de la variable de signal pour une observation i
- Prédiction(Yi) = valeur prédite de la variable de signal pour une observation i
- Moyenne(RésiduY, Estimation) : moyenne des valeurs des résidus pour le jeu de données d'estimation
- |Résidu(Yi) – Moyenne(RésiduY, Estimation)| : écart absolu entre les termes Résidu(Yi) et Moyenne(RésiduY, Estimation)|
- ÉcartType (Résidu(Y), Estimation) : écart standard des valeurs des résidus pour le jeu de données d'estimation

5.7.4.3 Affichage des prévisions

Le graphique des *Prévisions* vous permet de visualiser cinq types d'informations :

- le *signal*,
- le *signal prévu*,
- la *tendance*,
- les *intervalles de prédiction*,
- les *valeurs anormales*.

Chaque élément graphique et sa signification sont expliqués dans le tableau suivant :

L'élément...	Symbolisé par...	Représente...
Signal	la courbe verte	l'information contenue dans le jeu de données d'entraînement
Signal prévu	la courbe bleue	le signal prédit par le modèle généré Les périodes pouvant être prévues avec un niveau de confiance accepté sont représentées avec des intervalles de prédiction. On parle alors de périodes de confiance et le nombre affiché est déterminé par la taille du jeu de données et le type de stratégie de partition appliqué lors de l'entraînement du modèle. Consultez les rubriques liées pour connaître les règles utilisées pour calculer le nombre de périodes de confiance affichées.
Tendance	la courbe rouge	la tendance du signal. Cette courbe n'est affichée que dans le cas d'une tendance polynomiale ou linéaire.
Intervalles de prédiction	la zone bleue entourant la fin de la courbe bleue	la zone d'erreur possible du signal prévu. Les intervalles de prédiction ne sont affichés que pour les prévisions. **ℹ Remarque** L'intervalle de prédiction à un horizon donné H est un intervalle centré sur la valeur prévue, dont la largeur est calculée de la façon suivante : 1,96 * 2* EQM (erreur quadratique moyenne). L'EQM est calculée selon la valeur réelle par rapport à la valeur prévue observée dans un ensemble de validation lors de la prévision sur l'horizon H. La valeur pondérée de 1,96 correspond à un niveau de confiance de 95 %.

L'élément...	Symbolisé par...	Représente...
Valeur anormale	un carré rouge	un point où la courbe prédictive est très éloignée de la courbe réelle. Note - une valeur de signal réelle est qualifiée d'anormale lorsque l'erreur de prévision correspondante est considérée comme anormale par rapport à la moyenne observée pour le jeu de données d'estimation. L'indicateur d'erreur de prévision est la différence absolue entre les valeurs réelle et prédite. Elle est également appelée le résidu. Le seuil anormal du résidu est défini comme étant égal à trois fois l'écart standard des valeurs résiduelles pour un jeu de données d'estimation (ou de validation).

Le signal s'affiche uniquement sur le graphique si les prévisions sont disponibles. Le modèle peut cependant avoir besoin de plusieurs valeurs pour pouvoir commencer à générer des prévisions. Les premières valeurs du signal pour lequel aucune prévision n'a été générée n'apparaissent donc pas sur le graphique. Tant que le signal d'origine est affiché, vous pouvez mesurer la précision du signal prédit par rapport au signal d'origine. Lorsque le signal d'origine s'arrête, les intervalles de prédiction vous permettent de mesurer le niveau de confiance du signal prédit. Les intervalles de prédiction ne s'affichent pas au-delà de la première prévision pour laquelle le modèle ne peut pas garantir la précision du signal prédit. En d'autres termes, lorsque le modèle ne peut pas évaluer si la prévision est correcte, il cesse d'afficher les intervalles de prédiction.

> **i** Remarque
>
> Si vous avez sélectionné la stratégie de partition *Séquentielle*, les intervalles de prédiction s'affichent également sur la zone de test du signal.

Informations associées

How are the Number of Confidence Periods Determined? [page 198]

5.7.4.3.1 How are the Number of Confidence Periods Determined?

Confidence periods are the forecasted periods that can be plotted with an accepted level of confidence by the model. Each period estimates the uncertainty range of a forecasted value in the future at the horizon H, that comes from the modeling approximation built on the past observations of the time series. This estimate

at horizon H is proportional to the mean of prediction errors measured on the validation dataset at the same horizon H. The multiplicative coefficient depends on the confidence level.

The number of confidence periods is determined by the following rules:

Training dataset condition	Rule
If the training dataset size is equal or less than 12 periods.	It is treated as a small dataset case and by default the number of confidence periods is set to 1.
If the partition strategy is automatic.	The number of confidence periods is initialized at `.8 * validation size`, where validation corresponds to 1/4 of the whole training dataset.

5.7.5 Paramètres du modèle

Toutes les informations nécessaires au codage des variables, la compression, le remplacement des variables manquantes, la fréquence des variables, les paramètres du modèle, etc., sont stockées dans un arbre des paramètres. Cliquez sur l'icône de la loupe ou cliquez deux fois sur le répertoire pour afficher les informations. La plupart des informations sur le modèle figurent dans le répertoire Protocoles.

Le nom du sous-répertoire dans l'arbre des paramètres et la valeur du paramètre apparaissent dans la partie droite de l'écran.

Pour une description détaillée des différents paramètres, reportez-vous à la documentation des fonctionnalités de SAP Predictive Analytics.

5.7.6 Options de rapports statistiques

Vous avez toutes ou une partie des options suivantes à votre disposition quand vous utilisez les rapports statistiques :

Voir le rapport sous forme de table : le rapport est affiché comme table avec des lignes d'en-tête qui permettent un tri.

Voir le rapport HTML : le rapport est affiché comme table avec des couleurs par défaut ou définies par l'utilisateur.

Voir le rapport sous forme de diagramme à barres : quand cette icône est visible, le rapport peut être affiché sous forme de diagramme à barres.

Voir le rapport sous forme de diagramme à secteurs : quand cette icône est visible, le rapport peut être affiché sous forme de diagramme à secteurs.

Voir le rapport sous forme de courbe : quand cette icône est visible, le rapport peut être affiché sous forme de courbe.

Copier la vue actuelle : seul le rapport affiché est copié et peut ensuite être collé dans un autre programme.

Copier toutes les vues : s'il existe plusieurs vues d'un même rapport, vous pouvez les copier toutes en même temps et les coller dans un autre programme.

Copier l'image : si vous affichez le rapport comme diagramme, vous pouvez copier cette image et la coller dans un autre programme.

Imprimer : le rapport affiché est imprimé.

Enregistrer la vue actuelle : seul le rapport affiché est enregistré.

Enregistrer toutes les vues : s'il existe plusieurs vues d'un même rapport, vous pouvez les enregistrer toutes en même temps à l'aide de ce bouton.

Changer l'orientation des barres : quand le rapport en cours est affiché sous la forme d'un diagramme à barres, cette option vous permet de modifier son orientation (d'horizontal à vertical et inversement).

Exporter sous format Excel : le rapport affiché est exporté dans Excel sans perte de la mise en page comme c'est le cas en utilisant la fonction copier/coller.

> **i** Remarque
>
> Cette fonctionnalité est compatible avec Excel 2002, 2003, XP et 2007.

5.7.7 Rapports

5.7.8 Personnalisation des feuilles de style de rapports

Vous pouvez personnaliser les rapports générés. La feuille de style par défaut, appelée *Feuille de style de rapport (défaut)*, ne peut être modifiée. Vous devez créer vos propres feuilles de styles pour changer les paramètres.

> **i** Remarque
>
> Pour créer, charger ou enregistrer une feuille de style, vous devez indiquer une source de données dans le panneau `Modifier les optionsRapport` avant d'ouvrir la fenêtre *Éditeur de feuille de style de rapport*.

Référentiel des feuilles de style

- Pour créer une feuille de style
  1. Dans le champ *Dossier*, cliquez sur le bouton *Parcourir*.
  2. Sélectionnez un dossier.
     Ce dossier sera celui de vos feuilles de style.
  3. Cliquez sur le bouton *Ajouter*.
     Une feuille de style a été créée.
  4. Cliquez sur le bouton *Modifier*.
     Le panneau *Éditeur de feuilles de style* s'affiche.
  5. Dans le champ *Nom de la feuille de style*, saisissez le nom de la nouvelle feuille de style.
     L'extension .krs est ajoutée automatiquement.

     > **i** Remarque
     >
     > Vous pouvez dupliquer une feuille de style en modifiant son nom. La feuille précédente n'est pas supprimée.

- Pour supprimer une feuille de style
  1. Sélectionnez une des feuilles de styles proposées.
  2. Cliquez sur le bouton *Supprimer*.

     > **i** Remarque
     >
     > La feuille de style n'est pas seulement supprimée de la liste mais également de la source de données.

Modifier les paramètres généraux

Pour modifier les paramètres généraux

Paramètres...	Options...	Remarque...
Couleur de fond	• choisir une couleur • rendre transparent	L'affichage d'une couleur de fond n'est possible qu'avec les formats PDF et HTML.
Modifier la configuration	• taille de la police • style de la police • couleur de la police • couleur de fond des textes • configuration de tableau	Cochez l'option *Rendu dynamique des modifications des options* ou cliquez sur *Appliquer* lors de la modification des paramètres pour visualiser le résultat.

Les paramètres sélectionnés s'appliqueront à la fois à l'assistant et aux rapports générés.

Modifier des sections du rapport

Pour modifier des sections du rapport

1. Définissez les propriétés de votre choix.
2. Cliquez sur *Enregistrer* pour valider.
   Une fenêtre s'ouvre indiquant que votre feuille de style a été enregistrée.
3. Cliquez sur *OK*.

Propriétés...	Fonctions...	Remarque...
Affiché comme	nom du label	
Type de vue	Choisissez entre tabulaire, HTML et graphique. La dernière option n'est disponible que si la section peut être affichée comme graphe.	
Type de diagramme	Sélectionnez un des types proposés.	Cette option n'est disponible que pour les sections du type graphe.
Changer l'orientation des barres	Cette option vous permet de choisir une orientation différente de celle par défaut pour une section de rapport.	
Trier par/Ordre de tri	Vous pouvez choisir la colonne à utiliser pour le tri et l'ordre de tri.	
Visibilité	Vous pouvez masquer une colonne d'une section ou même toute une section de rapport.	Au moins une colonne d'une section doit rester visible.

5.8 Exécution du modèle

5.8.1 Application d'un modèle

Cette section décrit comment appliquer un modèle pour les différents types de modèles et fournir des informations sur les options de l'application.

5.8.1.1 Définition des paramètres pour l'application du modèle

Cet écran permet à l'utilisateur d'appliquer le modèle à un nouveau jeu de données ou au jeu de données utilisé pour le générer.

- Application d'un modèle [page 203]
- Options de génération [page 203]

- Mode de génération [page 209]

5.8.1.1.1 Application d'un modèle

Pour appliquer un modèle

1. Dans la section *Jeu de données d'application*, cliquez sur le bouton *Parcourir* sur la ligne *Dossier* pour sélectionner le sous-répertoire du fichier source.
2. Cliquez sur le bouton *Parcourir* sur la ligne *Données* pour sélectionner le fichier texte ou la source ODBC. Le format des données d'entrée doit être identique à celui utilisé pour créer le modèle (y compris la variable cible, même si celle-ci n'est pas renseignée).
3. Dans la section *Options de génération*, sélectionnez les résultats à générer dans la liste *Générer*. Les résultats dépendent du type de modèle appliqué.
4. Dans la section *Options de génération*, sélectionnez le mode de génération à utiliser dans la liste déroulante *Mode*.
5. Dans la section *Résultats générés par le modèle*, spécifiez le répertoire qui contiendra les résultats en cliquant sur le bouton *Parcourir* à côté de la ligne *Dossier*.
6. Saisissez un nom pour le fichier de sortie dans le champ *Données*, ou utilisez le bouton *Parcourir* pour sélectionner un fichier.
7. Pour afficher le résultat de la génération directement dans Automated Analytics, cochez l'option *Visualiser les données générées*.
8. Cliquez sur le bouton *Appliquer* une fois toutes les sélections effectuées.

5.8.1.1.2 Options de génération

En fonction de votre sélection dans la liste *Générer*, les données générées contiendront (en plus de toutes les variables clés) les éléments ci-dessous :

Classification/Régression

Si vous choisissez l'option...	les données générées contiendront dans chaque ligne...
Valeur prédite	Uniquement la valeur prédite des observations (rr_TargetVariableName)
Probabilité	• la valeur prédite • la probabilité (proba_rr_TargetVariableName) • les intervalles de prédiction (bar_rr_TargetVariableName)

Si vous choisissez l'option...	les données générées contiendront dans chaque ligne...
Contributions individuelles	• la valeur prédite • la probabilité • l'intervalle de prédiction • les contributions des variables (contrib_Variable-Name_rr_TargetVariableName)
Décision	• la valeur prédite • la décision (decision_rr_TargetVariableName) • la probabilité de la décision (proba_decision_rr_Target-VariableName) • la probabilité En sélectionnant l'option Décision dans la liste déroulante Générer, vous pouvez spécifier le nombre d'observations que le modèle doit détecter après l'application au nouveau jeu de données.
Paramètres avancés	Cette option vous permet de sélectionner les données de sortie qui figureront dans le fichier de résultats.

Mise en cluster

Si vous choisissez l'option...	les données générées contiendront dans chaque ligne...
Valeur prédite	• la valeur de la cible • le numéro du cluster contenant l'observation actuelle
Codage disjonctif des numéros de clusters	• la valeur de la cible • le numéro du cluster contenant l'observation actuelle le codage disjonctif des numéros de clusters, c'est-à-dire que pour chaque cluster une variable booléenne indiquant si l'observation appartient à ce cluster (valeur 1) ou non (valeur 0) est créée.
Codage disjonctif et recopie des var. explicatives	• la valeur de la cible • le numéro du cluster contenant l'observation actuelle le codage disjonctif des numéros de clusters (voir ci-dessus) c'est-à-dire que pour chaque cluster une variable booléenne indiquant si l'observation appartient à ce cluster (valeur 1) ou non (valeur 0) est créée. • toutes les variables d'entrée contenues dans le jeu de données d'application

Si vous choisissez l'option...	les données générées contiendront dans chaque ligne...
Moyenne de la cible pour les clusters	• la valeur de la cible • le numéro du cluster contenant l'observation actuelle • la moyenne de la cible pour le cluster contenant l'observation actuelle

Codage textuel

Si vous choisissez l'option...	les données générées contiendront dans chaque ligne...
Codage vectoriel	Toutes les colonnes présentes dans le jeu de données d'origine. Et pour chaque champ textuel : • une colonne est créée pour chaque racine identifiée par le modèle. Si la racine est présente dans l'enregistrement, la valeur de la colonne est 1, sinon elle est de 0. • une colonne indique le nombre d'éléments reconnus par l'encodage textuel dans cet enregistrement, • une colonne indique le nombre de racines distinctes trouvées dans l'enregistrement.
Reconnaissance de la langue	Pour chaque champ textuel : • une colonne indique la langue reconnue par l'encodage textuel pour cet enregistrement. Elle contient le code ISO de la langue ou est vide si aucune langue n'est reconnue.
Mode transactionnel	• une colonne pour chaque racine identifiée • une colonne contenant la clé de l'enregistrement original • une colonne contenant l'indice de la racine actuelle dans le champ textuel • une colonne indiquant de quel champ textuel la racine est extraite • une colonne contenant la racine actuelle Ce mode transforme chaque enregistrement en une série de transactions, créant ainsi un jeu de données utilisable dans l'analyseur de séquences.

Si vous choisissez l'option...	les données générées contiendront dans chaque ligne...
Générer seulement les racines	Pour chaque champ textuel : • une colonne est créée pour chaque racine identifiée par le modèle. Si la racine est présente dans l'enregistrement, la valeur de la colonne est 1, sinon elle est de 0. • une colonne indique le nombre d'éléments reconnus par l'encodage textuel dans cet enregistrement, • une colonne indique le nombre de racines distinctes trouvées dans l'enregistrement.

Pour un modèle Social

Options de génération

Si vous choisissez l'option...	les données générées contiendront dans chaque nœud...
Mode par défaut	• toutes les informations disponibles au niveau du nœud et de ses voisins
Mode cercle	• le nombre de voisins • des informations supplémentaires sur les voisins si elles sont disponibles • la moyenne des attributs des voisins
Mode centralité	• une évaluation de la centralité du nœud en analysant sa mise en cluster locale et le nombre de ses voisins
Mode voisinage	• une liste de tous les voisins du nœud et des informations supplémentaires sur eux
Mode description	• la liste des informations disponibles pour ce nœud
Mode communauté	• des informations sur la communauté à laquelle le nœud appartient et sur son rôle
Mode appariement	• des informations sur sa présences dans les graphes et sur ses voisins (nombre de voisins, rapport common_neighbors/average_neighbors_count)
Paramètres avancés	• des informations sélectionnées par l'utilisateur (classe du nœud, nombre de triangles, informations sur les voisins, index de la communauté, rôle du nœud dans la communauté, etc.).

Remarque sur le jeu de données d'application

Pour pouvoir appliquer un modèle social sur un jeu de données, celui-ci doit contenir les variables suivantes :

- une variable pour chaque population, par exemple une pour les clients et une pour les produits.
- la variable *kxComIndex* contenant les identifiants de communautés. Il doit s'agir d'un entier nominal.

Cette variable vous permet d'appliquer le modèle sur les graphes de communautés (détection des communautés voisines, liste des nœuds d'une communauté, statistiques agrégées, ...) Vous pouvez utiliser cette colonne pour fournir une liste d'identifiants de communautés pour lesquelles vous souhaitez effectuer des calculs. Si vous n'appliquez pas le modèle sur un graphe de communauté, cette variable ne sera pas utilisée, vous pouvez donc laissez cette colonne vide ou, dans le cas d'une base de données, utiliser une valeur factice.

Utilisez la fonctionnalité de correspondance disponible dans le panneau *Appliquer le modèle* pour garantir la cohérence des jeux de données.

Modèle de règles d'association

Si vous choisissez l'option...	les données générées contiendront dans chaque ligne...
Valeur prédite	les données de sortie simplifiées : • l'identifiant de session, • l'identifiant de la règle utilisée pour trouver le conséquent • le conséquent
Conséquents optimisés par KI	les données de sortie simplifiées Si plusieurs règles indiquent le même conséquent pour une session, la règle présentant la meilleure capacité prédictive (KI) sera sélectionnée.
Conséquents optimisés par confiance	les données de sortie simplifiées Si plus d'une règle indiquent le même conséquent pour une session, la règle pour laquelle l'indice de confiance est le plus élevé sera sélectionnée.
Valeur prédite avec description complète des règles	les données de sortie complètes • l'identifiant de session, • l'identifiant de la règle utilisée pour trouver le conséquent • le conséquent • l'antécédent • la capacité prédictive (KI) • l'indice de confiance • le support de la règle
Conséquents optimisés par KI avec description complète des règles	les données de sortie complètes Si plusieurs règles indiquent le même conséquent pour une session, la règle présentant la meilleure capacité prédictive (KI) sera sélectionnée.

Si vous choisissez l'option...	les données générées contiendront dans chaque ligne...
Conséquents optimisés par confiance avec description complète des règles	les données de sortie complètes Si plus d'une règle indiquent le même conséquent pour une session, la règle pour laquelle l'indice de confiance est le plus élevé sera sélectionnée.

Modèle de série temporelle

Si vous sélectionnez l'option...	Le fichier de résultat contiendra...
Valeur prédite	• la variable cible, c'est-à-dire la variable de signal • la variable de date • les variables prévues, c'est-à-dire les prévisions pour chaque date du jeu de données d'entraînement
les prévisions avec leurs composantes	• la variable cible, c'est-à-dire la variable de signal • la variable de date • les variables prévues, c'est-à-dire les prévisions pour chaque date du jeu de données d'entraînement • la valeur de chaque composante (tendance, cycles, fluctuation) pour chaque prévision
les prévisions avec leurs composantes et leurs résidus	• la variable cible, c'est-à-dire la variable de signal • la variable de date • les variables prévues, c'est-à-dire les prévisions pour chaque date du jeu de données d'entraînement • la valeur de chaque composante (tendance, cycles, fluctuation) pour chaque prévision • les valeurs résiduelles obtenues après extraction des composantes pour chaque prévision
uniquement la première colonne des prévisions avec leurs intervalles de prédiction	• la variable cible, c'est-à-dire la variable de signal • la variable de date • la première valeur prévue, c'est-à-dire la première prévision pour chaque date du jeu de données d'entraînement • les intervalles de prédiction pour la valeur prédite

5.8.1.1.3 Mode Application in-database

Ce mode optimisé permet d'appliquer le modèle directement dans la base de données. Cela évite d'avoir à extraire les données de la base de données et accélère le processus d'écriture des données de sortie du modèle.

Le mode Application in-database est disponible pour :

- Modèles de classification/régression avec un ordre du polynôme égal à 1.
- Modèles de segmentation générés avec des expressions SQL
- Recommandation et modèles sociaux

Vous pouvez utiliser ce mode si :

- Le jeu de données d'application d'entrée (table, vue, requête, manipulation de données) et le jeu de données de résultats (tables) provient de la même base de données.
- Au moins une variable du jeu de données d'entraînement utilisé pour créer le modèle est déclarée comme clé dans Automated Analytics.

  > **i** Remarque
  >
  > La (Les) clé (s) définie(s) dans le modèle est (sont) identifiée(s) en tant que clé primaire dans la table de sortie.

- L'utilisateur Automated Analytics qui déclenche la tâche d'application dispose des droits de lecture de l'entrée d'application et d'écriture de sortie d'application dans la base de données.
- Il existe un générateur d'indicateurs pour la base de données.
- Recommandation et modèles sociaux uniquement : Le modèle doit être enregistré dans la base de données contenant les jeux de données d'application.

  > **i** Remarque
  >
  > Si le modèle n'a pas été enregistré, une boîte de dialogue s'affiche et vous propose de l'enregistrer. Si vous n'enregistrez pas le modèle, Automated Analytics passe automatiquement au processus d'application standard.

> **i** Remarque
>
> Lorsque vous cochez l'option *Utiliser l'application directe dans la base de données*, l'option *Ajouter un écart de score* est automatiquement sélectionnée. Vous pouvez ainsi vérifier, pour chaque variable et chaque catégorie de variable, les écarts entre le modèle et le jeu de données d'entrée utilisé pour l'application du modèle.

5.8.1.1.4 Mode de génération

Selon l'option choisie dans la liste déroulante *Mode*, l'application :

- générera un fichier de sortie contenant toutes les données,
- générera un fichier de sortie ne contenant que les observations anormales (uniquement pour une régression),
- mettra à jour les statistiques sans générer un nouveau fichier de sortie.

En marquant toutes les variables comme clé, l'utilisateur pourra reproduire le jeu de données d'entrée avec les valeurs prédites, les probabilités, les intervalles de prédiction et les contributions.

5.8.1.2 Informations relatives au processus d'application

L'écran *Application du modèle* affiche des informations relatives au processus d'application à l'aide d'une barre de progression ou d'un journal. La barre de progression s'affiche par défaut. Pour voir le journal, cliquez sur le bouton *Afficher le détail du processus* situé dans la barre d'outils sous le titre du panneau.

Selon les paramètres sélectionnés pour l'application, les indicateurs listés ci-dessous peuvent apparaître dans le journal détaillé :

- *Indicateur de qualité capacité prédictive* (KI) : entre 0 (le modèle n'a extrait aucune information) et 1 (le modèle a parfaitement modélisé l'attribut cible). La qualité peut être améliorée en incluant d'autres variables.
  Indicateur de qualité dans le jeu de données d'application : KI=0.656589
- *Indicateurs standards d'erreur* : L1 (moyenne de la norme des erreurs), L2 (erreur quadratique - racine de l'erreur quadratique moyenne), LInf (erreur maximale), meilleur taux de classement.
  Erreurs standard dans le jeu de données d'application : L1=0.237272 L2=0.355389 LInf=1.11332
- *Temps total écoulé* : durée totale du processus d'application.
  Temps total écoulé : 16 secondes.

5.8.1.3 Décision de classification

Le panneau *Décision de classification* vous permet de choisir le seuil de décision pour chaque cible binaire avant d'appliquer le modèle en mode "Décision". Cette décision figurera pour chaque observation dans le fichier de résultats de l'application du modèle.

La décision repose sur l'un des critères de décision suivants :

- *Score :* indique le score utilisé pour séparer les observations positives des observations négatives.
- *Pourcentage attendu de cibles détectées* : détecte x% de la population cible.
- *Pourcentage attendu de la population visée :* détecte x% de la population totale.

Définitions

- On appelle *observation positive* toute observation *appartenant* à la *population cible*.
- On appelle *observation négative* toute observation n'appartenant pas à la *population cible*.

Paramètres

Cible

Sélectionnez la variable cible pour laquelle vous voulez afficher la matrice de confusion en cliquant sur l'onglet correspondant. Seules les cibles nominales sont disponibles.

Matrice de confusion

Il existe trois façons de définir le seuil à l'aide du curseur affiché :

- en sélectionnant le *pourcentage de population visé* si la population est triée par ordre décroissant de score (*% de la population*)
- en sélectionnant le *pourcentage d'observations positives* que vous souhaitez détecter (*% de cible détectée*)
- en sélectionnant directement le *score* à utiliser comme seuil (*Score*). Toute observation dont le score est supérieur au seuil est considérée comme positive et toute observation dont le score est inférieur au seuil est considérée comme négative.

L'échelle est graduée du plus petit score (à gauche), au plus grand (à droite). Les valeurs correspondant à chaque option sont affichées dans des champs situés sous l'échelle.

Lorsque vous déplacez le curseur sur l'échelle, la matrice de confusion est modifiée en conséquence. Le tableau ci-dessous indique comment lire la matrice de confusion.

	`Predicted[Target Category]` Observations positives prédites	`Predicted[Non-target Category]` Observations négatives prédites
`True[Target Category]` Observations positives réelles	Nombre d'observations positives correctement prévues	Nombre d'observations réellement positives mais prédites négatives
`True[Non-target Category]` Observations négatives réelles	Nombre d'observations réellement négatives mais prédites positives	Nombre d'observations négatives correctement prévues

Par défaut, la *Population totale* est égale au nombre d'enregistrements dans le jeu de données de validation. Vous pouvez modifier ce nombre pour visualiser la matrice sur la population sur laquelle vous voulez appliquer votre modèle.

Les Métriques

- Le *Taux de classification* correspond à la proportion de données correctement classée par le modèle lors de son application sur le jeu de données d'entraînement.
- La *Sensibilité* d'un test mesure sa capacité à donner un résultat positif lorsqu'une hypothèse est vérifiée.
- La *Spécificité* d'un test mesure sa capacité à donner un résultat négatif lorsque l'hypothèse n'est pas vérifiée.
- La *Précision* est le pourcentage avec lequel des mesures répétées dans des conditions inchangées donnent des résultats identiques.
- Le *Score* indique à quel point la fonction de vraisemblance dépend de son paramètre.
  La fonction de vraisemblance est une fonction de probabilités conditionnelles qui décrit les valeurs d'une loi statistique en fonction de paramètres supposés connus.

Profit

Cette section vous permet de visualiser votre profit par rapport au score sélectionné, ou bien de sélectionner automatiquement le score d'après les paramètres de profit.

Pour chaque catégorie d'observations, saisissez un profit ou un coût par observation. Le profit total est automatiquement affiché dans le champ Profit situé à droite.

Pour sélectionner automatiquement le seuil de score qui vous permettra de faire le maximum de profit, remplissez la matrice de coût puis cliquez sur le bouton *Maximiser le profit*.

Exemple

Dans le tableau profit-coût ci-dessous, chaque observation positive correctement identifiée rapportera 15 € mais chaque observation négative identifiée comme positive coûtera 8 €.

Catégorie	Prévu[1]	Prévu[0]
Réel[1]	15	0
Réel[0]	-8	0

5.8.1.4 Visualisation du code source

Ce panneau affiche le code généré qui a été enregistré. Vous pouvez ainsi vérifier le code avant de le déployer et si nécessaire le re-générer.

5.8.1.5 Application d'un modèle de série temporelle

Pour créer le graphique des prévisions, le module Série temporelle applique le modèle au jeu de données d'entraînement et génère les prévisions.

Une barre de progression vous permet de suivre les différentes étapes du processus. Une fois le modèle appliqué, le graphique *Voir les prévisions* est affiché.

5.8.2 Simulation du modèle

Le modèle en cours d'utilisation peut être utilisé pour effectuer des simulations sur des observations spécifiques, au cas par cas. Pour définir l'observation à analyser, vous renseignez les variables de votre choix. Lors de la simulation, SAP Predictive Analytics affectera automatiquement les valeurs manquantes essentielles à la bonne exécution de la simulation.

Suite à la simulation, vous obtenez les résultats suivants :

- la **valeur prédite** (score),

- la **probabilité** de cette observation d'appartenir à la catégorie cible de la variable cible.

- Pour simuler un modèle
  1. Sur l'écran *Utilisation du modèle*, cliquez sur l'option *Simulation*.
     L'écran *Simulation du modèle* s'affiche.
  2. Sur la partie gauche de l'écran (*Variables explicatives*), sélectionnez une variable.
     Ses valeurs apparaitront dans la section *Modification des valeurs* à droite de l'écran.
  3. Dans la section *Modification des valeurs*, dans le champ *Valeur*, sélectionnez ou saisissez une valeur.
     La valeur apparaitra dans la table *Variables explicatives* en face de la variable sélectionnée.
  4. Si vous souhaitez sélectionner d'autres variables explicatives, retournez à l'étape 2.
     Sinon, passez à l'étape 5.
  5. Cliquez sur le bouton *Exécuter* pour effectuer une simulation du modèle.
     Les résultats de la simulation apparaissent dans la partie *Résultat*. Vous obtenez la **valeur prédite** (score) de l'observation décrite dans la table *Variables explicatives* ainsi que la **probabilité** que cette observation appartienne à la catégorie cible de la variable cible.
     Il est à noter que certaines variables de la table *Variables explicatives* ont été complétées automatiquement lors de l'exécution de la simulation. Le modèle a en réalité ajouté automatiquement certaines valeurs manquantes qui étaient essentielles à la simulation.
     Les valeurs sont indiquées dans le tableau ci-dessous.

Type de variable	Valeur par défaut
Variable continue	Valeur moyenne
Variable nominale	la catégorie la plus fréquente
Valeur ordinale	la catégorie la plus fréquente

     Ces modifications apparaissent dans la partie gauche de l'écran lorsque vous cliquez sur le bouton *Exécuter*.

  6. Vous pouvez modifier la valeur d'une variable explicative et exécuter à nouveau la simulation pour mesurer l'impact d'un tel changement par rapport à la variable cible.
  7. Cliquez sur le bouton *Réinitialiser* pour effectuer une nouvelle simulation du modèle.

5.8.3 Analyse des écarts

Le panneau *Jeu de données utilisé pour la vérification des écarts* vous permet de sélectionner le jeu de données pour lequel vous voulez vérifier s'il y a des écarts par rapport au jeu de données utilisé pour la création du modèle.

- Pour sélectionner un jeu de données
  1. Sélectionnez le type de fichier de jeu de données dans la liste *Type de données*.
  2. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de données contenant le jeu de données.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Données* pour sélectionner le fichier ou la table contenant le jeu de données.
  4. Cliquez sur le bouton *Suivant*.

- Options avancées
  - Cliquez sur le bouton pour modifier les paramètres avancés (échantillonnage, paramètres des fichiers, définition de la ligne d'en-tête).
  - Cliquez sur le bouton pour afficher le jeu de données ou les statistiques du jeu de données.
  - Cliquez sur le bouton *Définir les correspondances* pour associer les variables trouvées dans le nouveau jeu de données à celles du jeu de données d'origine.

5.8.4 Analyse des écarts

Ce panneau vous permet de suivre le processus de test à l'aide d'une barre de progression. A la fin du processus, un panneau récapitulatif s'affiche.

Vous pouvez utiliser la barre d'outils située en haut du panneau pour :

- afficher le journal détaillant l'historique du processus de vérification,
- copier, imprimer ou enregistrer le récapitulatif de la vérification.

Aperçu

Jeu de données	Nom du jeu de données d'application
Source	Stockage du jeu de données
Nombre d'enregistrements	Nombre d'enregistrements dans le jeu de données
Nombre de variables montrant des écarts	Nombre de variables pour lesquelles des écarts ont été trouvés par rapport au jeu de données d'origine utilisé pour entrainer le modèle

Variables déviantes

Variable	Nom de la variable pour laquelle un écart a été détecté par rapport au jeu de données d'origine utilisé pour entrainer le modèle
Probabilité d'écart	Valeur entre 0 et 1 qui indique l'écart d'une variable d'un jeu de données à l'autre. Une probabilité d'écart de 1 indique qu'aucune valeur correspondante de cette variable n'a été trouvée dans les deux jeux de données.

Indicateurs de performance

Pour chaque cible

rr_<TargetName>	Nom de la cible
kc_<TargetName>	**i Remarque** rr_ indique une régression robuste (K2R) et kc_ une mise en cluster (K2S).
KI	Capacité prédictive, indicateur de qualité qui correspond à la proportion d'information contenue dans la variable cible que les variables explicatives peuvent expliquer.
KR	Confiance prédictive, indicateur de robustesse qui précise la capacité du modèle à obtenir les mêmes performances lorsqu'il est appliqué à un nouveau jeu de données ayant les mêmes caractéristiques que le jeu de données d'entraînement.

Performance sur le jeu de données de contrôle

La section *Performance sur le jeu de données de contrôle* affiche pour chaque variable cible les indicateurs *KI* et *KR* obtenus par le modèle sur le jeu de données de contrôle.

Si le KI et/ou le KR du modèle sur le jeu de données de contrôle sont significativement *plus faibles*, cela signifie que la relation entre les variables et la variable cible a changé, et en conséquence, que le *modèle devrait être généré à nouveau* sur les nouvelles données.

Si le *KI* et le *KR* n'ont pas ou peu changé, cela signifie que la relation entre les variables d'entrée et la variable cible est toujours la même, mais cela ne signifie pas qu'il n'y a aucune différence de distribution entre les jeux de données.

5.8.5 Sélection des variables contributives

Ce panneau vous permet d'affiner le modèle actuellement ouvert. Par exemple, vous pouvez :

- réduire le nombre de variables explicatives utilisées par le modèle tout en conservant la qualité (capacité prédictive) et la robustesse (reproductibilité) initiales ;
- sélectionner les variables les plus importantes à l'ordre 1 pour construire un modèle d'ordre 2.

Si certaines variables ont été déclarées comme variables de date (date ou date et heure), des variables supplémentaires créées par le composant de codeur de date apparaissent dans la liste. Il extrait automatiquement des variables de dates des informations spécifiques telles que l'année, le jour de la semaine, ou le trimestre et crée les variables correspondantes qui sont alors utilisées comme variables d'entrée du modèle.

- Sélection automatique des variables [page 216]
- Sélection des variables corrélées [page 217]

5.8.5.1 Sélection automatique des variables

La *Sélection intelligente* vous permet de laisser l'application choisir les variables ayant les plus fortes contributions selon la quantité d'information que vous souhaitez conserver.

Pour chaque variable, les informations suivantes sont fournies :

- L'index de la variable (*Index*)
- Le nom de la variable (*Variable*)
- La contribution maximale de la variable (*Max Contribution*)
- La capacité prédictive individuelle de la variable (*KI*), qui représente la capacité de cette variable seule de prédire la variable cible
- La reproductibilité individuelle de la variable (*KR*)
- Les corrélations trouvées pour cette variable (*r*). Si d'autres variables sont corrélées à cette variable, l'indicateur de corrélations est allumé.

Par défaut, les variables sont triées par contributions maximales décroissantes.

- Pour utiliser la Sélection intelligente :
  1. Dans la liste déroulante *Cibles*, sélectionnez la variable pour laquelle vous voulez affiner le modèle.
  2. Cliquez sur le bouton *Sélection intelligente*. La fenêtre Sélection intelligente affiche le pourcentage cumulé de contribution à conserver.
  3. Déplacez le curseur sur l'échelle pour modifier le *pourcentage d'information conservée*. Le nombre de variables conservées comme variables d'entrée est automatiquement modifié. Les variables conservées sont celles ayant les plus fortes "contributions intelligentes".
     Notez que les variables d'entrée n'apportant aucune information (telles que les constantes) sont automatiquement exclues par le modèle durant la phase d'apprentissage. Un rappel de ces variables exclues est affiché sous l'échelle.
  4. Une fois la quantité d'information sélectionnée, cliquez sur le bouton **OK**. La fenêtre *Sélection intelligente des variables* se ferme et les variables à conserver sont automatiquement sélectionnées.
  5. Cliquez sur le bouton *Suivant*. Une boîte de dialogue s'affiche vous demandant de confirmer que vous voulez annuler le modèle en cours pour le relancer avec les variables sélectionnées.
  6. Si les données sont stockées en base de données, la fenêtre *Connexion à la source de données* s'ouvre.
     1. Utilisez les champs *Nom* et *Mot de passe* pour saisir les informations de connexion correspondant à la base de données.
     2. **Vérifiez** que le *référentiel de métadonnées* défini dans vos options est identique à celui utilisé pour créer le modèle.

     Si ce n'est pas le cas, cliquez sur le bouton et sélectionnez le référentiel de métadonnées ayant servi à créer le modèle. Le référentiel de métadonnées défini ici ne sera utilisé que pour cette opération. Le référentiel de métadonnées défini dans les paramètres ne sera pas modifié.
  7. Cliquez sur le bouton *OK*. Le panneau *Sélection des variables explicatives* s'affiche et les variables qui n'ont pas été sélectionnées lors de l'étape précédente apparaissent dans la liste des Variables exclues.
  8. Vous pouvez alors relancer l'apprentissage du modèle sur ces variables.

5.8.5.2 Sélection des variables corrélées

Dire que des variables sont corrélées signifie qu'elles sont en partie redondantes, qu'elles apportent partiellement la même information par rapport à la variable cible. Deux variables dites hautement corrélées décriraient la même information, ou le même concept, à un degré encore plus élevé.

Quand deux variables A et B sont fortement corrélées :

- La variable A, avec une contribution supérieure à la variable B par rapport à la variable cible, devient la "variable primaire" : elle apparait en premier dans la liste des variables.
- La variable B, avec une contribution inférieure à la variable A par rapport à la variable cible, devient la "variable secondaire" : seule sa contribution marginale est affichée dans la colonne *Contribution max*.

Cependant, plusieurs raisons, telles que la difficulté d'obtenir certaines variables ou le prix d'achat des données, peuvent vous pousser à choisir quelle variable doit être considérée comme la variable primaire. Une fois que les variables les plus importantes ont été sélectionnées à l'aide de la **sélection intelligente**, vous pouvez choisir parmi les variables corrélées à celles sélectionnées celles à conserver.

- Pour sélectionner les variables corrélées à utiliser :
  1. Dans la colonne **r**, cliquez deux fois sur l'icône colorée correspondant à la variable. La fenêtre *Corrélations* s'ouvre affichant une liste des variables corrélées à la variable en cours. La variable en cours est visiblement répertoriée avec un coefficient de corrélation de 1.0, les autres variables corrélées à cette variable sont listées par ordre décroissant du coefficient de corrélation.
  2. Si vous préférez utiliser une autre variable que celle en cours, cochez la case *Sélection* correspondant à cette variable et décochez celle pour la variable en cours.
  3. Vous pouvez également sélectionner plusieurs variables mais privilégier l'utilisation de l'une par rapport à l'autre lors de la modélisation en cochant la case *Priorité augmentée*.

> **i** Remarque
>
> La priorité augmentée est signalée dans la liste des variables par un + rouge sur l'icône de corrélation.

5.9 Enregistrement/Export

5.9.1 Génération du code

Cet écran vous permet de générer un fichier de code compilable à partir du modèle créé. Le modèle créé sera identique au modèle généré par SAP Predictive Analytics.

- Paramètres standard [page 218]
- Paramètres avancés [page 220]

5.9.1.1 Paramètres standard

Le fichier de code généré par SAP Predictive Analytics contiendra toutes les informations nécessaires pour le modèle, comme le codage des variables, les valeurs de remplacement des valeurs manquantes, les compressions et les paramètres du modèle.

- Pour générer le code correspondant au modèle
  1. Dans la liste *Cible à utiliser*, sélectionnez la cible du modèle.
  2. Dans la section *Options de génération*, sélectionnez l'option que vous souhaitez :

Option sélectionnée	Résultats du modèle généré
Score/Estimations	le score (classification) ou les estimations (régression)
Probabilité	le score et la probabilité (sauf pour HTML et tous les codes SQL, pour lesquels seule la probabilité est donnée)
Bar	le score et l'intervalle de prédiction (sauf pour HTML et tous les codes SQL, pour lesquels seul l'intervalle de prédiction est donné)

> ⚠ Attention
>
> Les deux options *Probabilité* et *Bar* sont disponibles uniquement pour les modèles de classification/régression avec des cibles nominales.

  3. Dans la liste *Paramétrage du code*, sélectionnez le type de code que vous voulez générer.
  4. Cliquez sur le bouton *Parcourir* associé au champ **Dossier** et sélectionnez le dossier où vous voulez enregistrer le fichier généré.
  5. Dans le champ *Fichier généré*, saisissez le nom du fichier exporté. Pour remplacer un fichier existant, sélectionnez-le à l'aide du bouton *Parcourir*.
  6. Si vous avez sélectionné l'option *Visualiser le code généré*, celui-ci s'affiche à la fin de la génération.

Codes générés

Liste des codes disponibles avec leurs particularités

Code généré	Commentaire
Code AWK	
Code C	se référer à la documentation du générateur de code C
Code CCL	pour la définition du score dans le streaming des données intelligentes SAP HANA. uniquement disponible pour des modèles de classification, régression et mise en cluster.
Code CSV	
Code PMML 3.2	
Code Cpp	

Code généré	Commentaire
DB2 UDF (SQL)	
HTML (Javascript)	contient un formulaire permettant de reproduire le modèle
Code JAVA	le fichier KxJRT.jar est nécessaire à sa compilation et son exécution
Oracle UDF (SQL)	
Code SAS	
Code SQL (ANSI)	
Code SQL pour Greenplum	
Code SQL pour Hive	
Code SQL pour IMB DB2	
Code SQL pour MySQL	
Code SQL pour Netezza	
Code SQL pour Oracle	
Code SQL pour PostgreSQL	
Code SQL pour SAP HANA	
Code SQL pour SQLServer	entoure les noms de variables avec [ ]
Code SQL pour Spark	
Code SQL pour Sybase IQ	
Code SQL pour Teradata	
Code SQL pour VERTICA	
SQLServer 2000 UDF (SQL)	
Scorecard	uniquement disponible pour des modèles de classification/régression
Teradata V2R5.1 UDF	
Code UDF pour MySQL	
Code UDF pour Sybase IQ	
Code VB	

i Remarque

Lors de la génération des codes indiqués dans le tableau ci-dessous, vous allez devoir fournir des informations complémentaires.

Code généré	Informations à fournir
• SQL • SAS • Code SQL pour MySQL	• nom de la colonne clé • nom du jeu de données utilisé

Code généré	Informations à fournir
• Code CCL	• nom de la colonne clé • nom du flux d'entrée utilisé/de la fenêtre utilisée

5.9.1.2 Paramètres avancés

Mode UNICODE

Le *Mode Unicode* vous permet de générer le code choisi en Unicode pour qu'il puisse prendre en charge les langues non-latines telles que le japonais, le russe, etc.

> **i** Remarque
>
> Cette option s'applique en particulier aux codes SQL.

Options SQL/UDF

- L'option *Ne pas générer le code pour les variables non contributives* vous permet d'exclure du code toutes les variables ayant une contribution de 0 puisqu'elles n'influencent pas le résultat. Dans certains cas, cette option peut réduire d'une façon significative la taille du code généré.
- Vous pouvez soit *Utiliser le séparateur par défaut* ("GO"), soit *Utiliser un séparateur personnalisé*.

5.9.2 Génération d'un script KxShell

Ce panneau vous permet de générer un script KxShell reproduisant le modèle en cours. Ce script peut être utilisé pour entraîner des modèles par lots.

Un moyen facile d'obtenir des paramètres spécifiques dans les scripts KxShell exportés est d'effectuer les opérations correspondantes dans l'interface graphique avant de générer le code. Par exemple, si vous faites une sélection automatique des variables avant l'export du script shell, celui-ci contiendra le code nécessaire à cette opération.

- Pour enregistrer le script KxShell
  1. Cliquez sur le bouton *Parcourir* situé à droite du champ **Fichier**.
  2. Dans la fenêtre *Choisir un fichier*, sélectionnez le dossier où vous souhaitez enregistrer le script.
     - Si le fichier existe déjà, sélectionnez-le.

- Sinon, entrez "/" puis le nom du nouveau fichier après le nom du dossier.

3. Cliquez sur le bouton *OK*.

- Pour exporter le script en vue de l'apprentissage du modèle
  1. Cochez l'option *Apprentissage*.
  2. Dans le cadre *Enregistrement de description des jeux de données modèles*, sélectionnez l'emplacement où vous voulez enregistrer la description des données. Les options suivantes sont proposées :
     - *Enregistrer les descriptions dans le script*
       La description des données est ajoutée dans le script KxShell. Un seul fichier est généré.
     - *Enregistrer les descriptions au même endroit que le script*
       La description des données est enregistrée dans un nouveau fichier situé dans le même dossier que le script KxShell.
     - *Enregistrer les descriptions au même endroit que les données*
       La description des données est enregistrée dans un nouveau fichier situé dans le même dossier que les données utilisées pour le modèle.
     - *Enregistrer les descriptions à part*
       La description des données est enregistrée dans un nouveau fichier. Indiquez le type de description (fichier texte, base de données, espace de stockage mémoire, etc.) et l'emplacement où la description doit être enregistrée.

     > **i Remarque**
     >
     > Lors de l'enregistrement de la description dans un fichier supplémentaire, le fichier est nommé selon la syntaxe suivante : KxDesc_<Rôle jeu de données>_<Nom jeu de données>. Par exemple, dans le cas d'un jeu de données d'entraînement appelé Census.csv, le fichier de la description s'appellera KxDesc_Apprentissage_Census.csv.

  3. Pour des modèles avec une variable cible, vous pouvez exporter la structure de la variable en lien avec la cible en cochant l'option **Exporter la structure de variable dans le script** et en sélectionnant la variable cible dans la liste *Sélectionner une cible*. Cette option vous permet de forcer le regroupement de catégories lors de l'entraînement du modèle avec de nouveaux jeux de données.
  4. Cliquez sur le bouton *Suivant* pour lancer la génération.
- Pour exporter le script en vue de l'application du modèle
  1. Cochez l'option *Appliquer*.

     > **i Remarque**
     >
     > Vous devez appliquer le modèle une fois dans l'assistant avant d'exporter le script.

  2. Sélectionnez la version du modèle que vous voulez utiliser.
     - *Utiliser la version la plus récente du modèle* : le script appliquera toujours la dernière version créée du modèle.
     - *Utiliser la version actuelle du modèle* : le script appliquera toujours la version utilisée pour l'exporter.
  3. Cliquez sur le bouton *Suivant* pour lancer la génération.

5.9.3 Enregistrement du modèle

Enregistrer le modèle L'enregistrement conserve la totalité des informations qui sont relatives au modèle, c'est-à-dire ses paramètres de modélisation, ses courbes de profits, etc.

Enregistrer le modèle

- Pour enregistrer un modèle
  1. Sur l'écran *Utilisation du modèle*, cliquez sur l'option *Enregistrer le modèle en cours*.
     L'écran *Enregistrement du modèle* s'affiche.
  2. Renseignez les champs suivants :
     - *Nom du modèle :* Ce champ vous permet d'associer un nom au modèle. Ce nom est utilisé dans la liste des modèles qui vous est proposée quand vous chargez un modèle existant.
     - *Description* : Ce champ vous permet de saisir les informations de votre choix telles que le nom du jeu de données d'entraînement utilisé, l'ordre du polynôme ou les indicateurs de performance (capacité prédictive (KI) et reproductibilité (KR)) obtenus. Ces informations peuvent vous être utiles ultérieurement pour identifier le modèle.
     - **Type de données** : Cette liste vous permet de sélectionner le format dans lequel votre modèle sera enregistré. Les formats suivants sont proposés :
       - *Fichiers texte*, pour enregistrer le modèle dans un fichier texte
       - *Bases de données*, pour enregistrer le modèle dans une base de données
       - *Espace de stockage mémoire*, pour enregistrer le modèle dans la mémoire active
       - *Fichiers SAS*, pour enregistrer le modèle dans un fichier compatible SAS pour une version indiquée de SAS et une plateforme indiquée (SAS v6 ou 7/8 pour Windows ou Linux)
       - *SAS Transport*, pour enregistrer le modèle dans un fichier compatible SAS générique
     - *Dossier* : En fonction de l'option que vous avez sélectionnée, ce champ vous permet de spécifier la source ODBC ou le dossier dans lequel vous souhaitez enregistrer le modèle.
     - *Fichier/Table* Ce champ vous permet d'entrer le nom du fichier ou de la table qui contiendra le modèle. Le nom de fichier doit contenir l'une des deux extensions de format **.txt** (fichier texte dans lequel les données sont séparées par des tabulations) ou **.csv** (fichier texte dans lequel les données sont séparées par des virgules).

Fichiers créés lors de l'enregistrement du modèle

Lorsque vous enregistrez un modèle, plusieurs fichiers sont créés à l'emplacement indiqué. Le tableau ci-dessous liste les fichiers et les tables créés et dans quel cas.

Nom du fichier	Description	Utilisé par
KxAdmin	liste tous les modèles contenus dans le dossier ou la base de données ainsi que leurs informations de base (date, version, nom du modèle, commentaires)	tous les modèles créés avec Automated Analytics
<Nom_modèle>	fichier nommé d'après le modèle et contenant toutes les données à l'exception des informations des graphes. Ces dernières sont stockées dans des tables ou fichiers supplémentaires (voir ci-dessous)	tous les modèles créés avec Automated Analytics
KxInfos	indique quelles tables additionnelles sont utilisées par le modèle	tous les modèles créés avec Automated Analytics
KxOlapCube	contient les informations du cube OLAP utilisé par l'arbre de décision, lorsque l'option Arbre de décision a été activée	modèles de classification/régression avec arbre de décision
KxLinks	contient les liens des graphes du modèle	modèles sociaux uniquement
KxNodes	liste l'ensemble des nœuds de tous les graphes et leurs attributs	modèles sociaux uniquement
KxCommunities	contient les correspondances entre les nœuds et leur communauté lorsque la détection des communautés a été activée	modèles sociaux uniquement

> **i Remarque**
>
> Lorsque vous partagez ou envoyez un modèle, vous devez y joindre tous ces fichiers, autrement le destinataire ne pourra pas ouvrir le modèle.

5.9.4 Enregistrement et exportation d'un jeu de données d'analyse

Ce panneau permet d'enregistrer le jeu de données d'analyse venant d'être créé dans votre base de données.

- Pour enregistrer le jeu de données d'analyse
  1. Saisissez le nom du nouveau jeu de données d'analyse dans le champ *Nom du jeu de données d'analyse*. C'est par ce nom que vous pourrez identifier le jeu de données dans votre base de données.

2. Saisissez une description du jeu de données dans le champ *Description*.
3. Cliquez sur le bouton *Enregistrer*. Un message indiquant *Manipulation de données enregistrée* s'affiche.
4. Cliquez sur le bouton *Annuler* pour revenir au menu principal.
5. Cliquez sur *Oui* dans la boîte de dialogue demandant si vous voulez annuler le modèle en cours (cela n'annulera pas la manipulation de données que vous venez d'enregistrer).

5.9.5 Sauvegarde automatique

Ce panneau vous permet d'activer la sauvegarde automatique et de définir les paramètres requis lors de l'enregistrement du modèle.

Paramètre	Description
Nom du modèle	Ce champ vous permet d'associer un nom au modèle. Ce nom est utilisé dans la liste des modèles qui vous est proposée quand vous chargez un modèle existant.
Description	Ce champ vous permet de saisir les informations de votre choix telles que le nom du jeu de données d'entraînement utilisé, l'ordre du polynôme ou les indicateurs de performance (capacité prédictive (KI) et reproductibilité (KR)) obtenus. Ces informations peuvent vous être utiles ultérieurement pour identifier le modèle. **i Remarque** Cette description remplacera celle saisie dans le panneau Récapitulatif des paramètres de modélisation.
Type de données	Cette liste vous permet de sélectionner le format dans lequel votre modèle sera enregistré. Les formats suivants sont proposés : • Fichiers texte, pour enregistrer le modèle dans un fichier texte • Bases de données, pour enregistrer le modèle dans une base de données • Espace de stockage mémoire, pour enregistrer le modèle dans la mémoire active • Fichiers SAS, pour enregistrer le modèle dans un fichier compatible SAS pour une version indiquée de SAS et une plateforme indiquée (SAS v6 ou 7/8 pour Windows ou Linux) • Transport SAS, pour enregistrer le modèle dans un fichier compatible SAS générique.
Dossier	En fonction de l'option que vous avez sélectionnée, ce champ vous permet de spécifier la source ODBC ou le dossier dans lequel vous souhaitez enregistrer le modèle.

Paramètre	Description
Fichier/Table	Ce champ vous permet d'entrer le nom du fichier ou de la table qui contiendra le modèle. Le nom du fichier doit contenir l'une des deux extensions de format .txt (fichier texte dans lequel les données sont séparées par des tabulations) ou .csv (fichier texte dans lequel les données sont séparées par des virgules).

5.10 Social

Cette section décrit les concepts sous-jacents à la création d'analyses de réseaux sociaux dans SAP Predictive Analytics et comment créer, exposer et utiliser ce type de modèle.

5.10.1 Concepts

5.10.1.1 Types de jeu de données

Il est possible d'utiliser trois types de jeux de données dans Social :

- Jeu de données de liens (obligatoire)
- Jeu de données d'attributs descriptifs (facultatif)
- Jeu de données de conversion d'identifiants (facultatif)

Informations associées

Jeu de données de liens [page 225]
Jeu de données d'attributs descriptifs [page 226]
Jeu de données de conversion des identifiants [page 226]

5.10.1.1.1 Jeu de données de liens

Ce jeu de données sert à définir l'emplacement des nœuds ainsi que les informations à utiliser pour créer des liens à partir de données brutes.

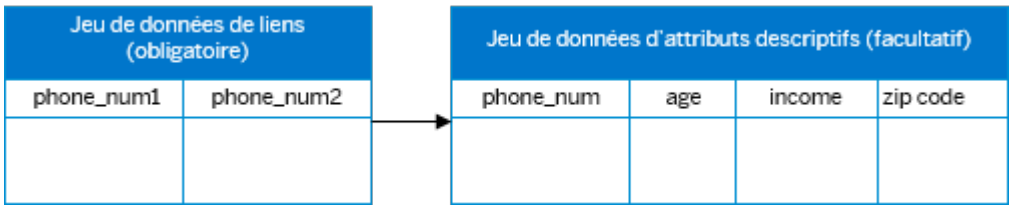
Une des fonctionnalités principales du composant est la capacité à générer des graphes à partir de différents types de données stockées dans des bases de données opérationnelles. Cela est réalisable en un seul nettoyage à partir de données brutes, en appliquant des filtres définis par l'utilisateur. Cela signifie que pour chaque cas du jeu de données d'entrée, l'application vérifie si la ligne peut être utilisée pour ajouter un lien dans l'un des graphes définis, ou si elle doit être ignorée.

5.10.1.1.2 Jeu de données d'attributs descriptifs

L'application permet d'ajouter des attributs descriptifs aux nœuds en associant des variables supplémentaires, telles que l'âge, l'état civil, le sexe ou la profession. Pour ce faire, il suffit de définir la colonne contenant ces informations qui seront extraites lors du processus de chargement.

Remarque : ces informations sont liées à l'aide des ID trouvés dans le jeu de données et ceux utilisés dans le graphe (qui se trouve dans le jeu de données de liens ou indirectement via le jeu de données de conversion d'identifiants).

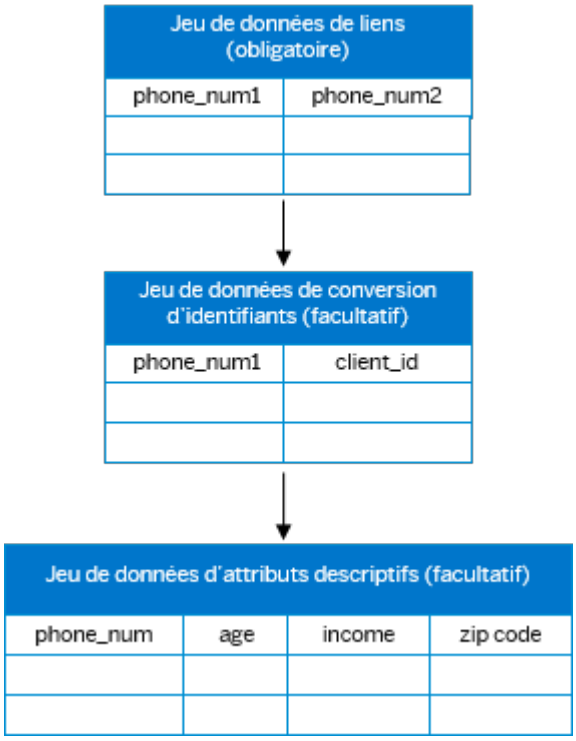
Si le jeu de données d'attributs descriptifs ne contient pas les mêmes ID que le jeu de données de liens, vous devez utiliser un jeu de données de conversion d'identifiants pour lier ces deux tables.



5.10.1.1.3 Jeu de données de conversion des identifiants

Il s'agit d'un jeu de données à deux colonnes employé pour mapper un ID depuis le jeu de données de liens vers un ID utilisé en tant que clé de nœud. Il est possible qu'un ID trouvé dans le jeu de données de liens ne soit pas celui attendu ou souhaité dans le graphe.

Par exemple, si un graphe est conçu à partir d'un enregistrement détaillé d'un appel (numéros de ligne téléphonique), vous souhaiterez peut-être le convertir automatiquement en ID client. C'est la raison pour laquelle l'application fournit une méthode adaptée à une telle conversion lors du processus de chargement. En outre, cette table permet de déduire les informations sur le réseau et hors ligne : un ID trouvé dans le jeu de données de conversion d'identifiants est considéré comme "sur le réseau". Sinon, il est considéré comme "hors réseau".



5.10.1.2 Types de graphes

Les sections suivantes présentent différents types de graphes utilisés dans l'application.

5.10.1.2.1 Contacts

Ce type de graphe décrit les contacts de communication entre le Nœud 1 et le Nœud 2 (il s'agit dans le cas présent de numéros de téléphone) :

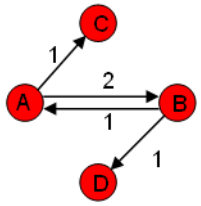
Nœud 1	Nœud 2	Type	Date	Durée
0143525691	0631527812	MESSAGE VOCAL	03/12/2007	0 :01 :12
0143525691	0631527812	MESSAGE VOCAL	03/13/2007	0 :04 :13
0499472604	0325363819	SMS	03/14/2007	0 :12 :52
0143525691	0631527812	MESSAGE VOCAL	03/15/2007	0 :05 :42
0186076517	0631527812	SMS	03/16/2007	0 :05 :09
0189340041	0154876556	MMS	03/17/2007	0 :36 :04

Il est possible d'utiliser la même condition de filtre que pour l'option Liens uniquement (voir page 14), mais une condition supplémentaire est disponible : les nœuds sont liés, si (et seulement si), un événement entre eux se produit au moins n fois (voir les exemples en bleu dans le tableau).

Exemples de graphes :

- Graphe de clients possédant au moins 3 messages vocaux.
- Graphe de clients possédant au moins 5 communications par SMS entre le 03/01/2007 et le 04/01/2007
- Graphe de clients possédant au moins 2 communications vocales de 120 secondes entre le 03/01/2007 et le 04/01/2007

Node 1	Node 2
A	B
A	C
B	D
A	B
B	A



(directed graph)

Informations associées

Liens uniquement [page 230]

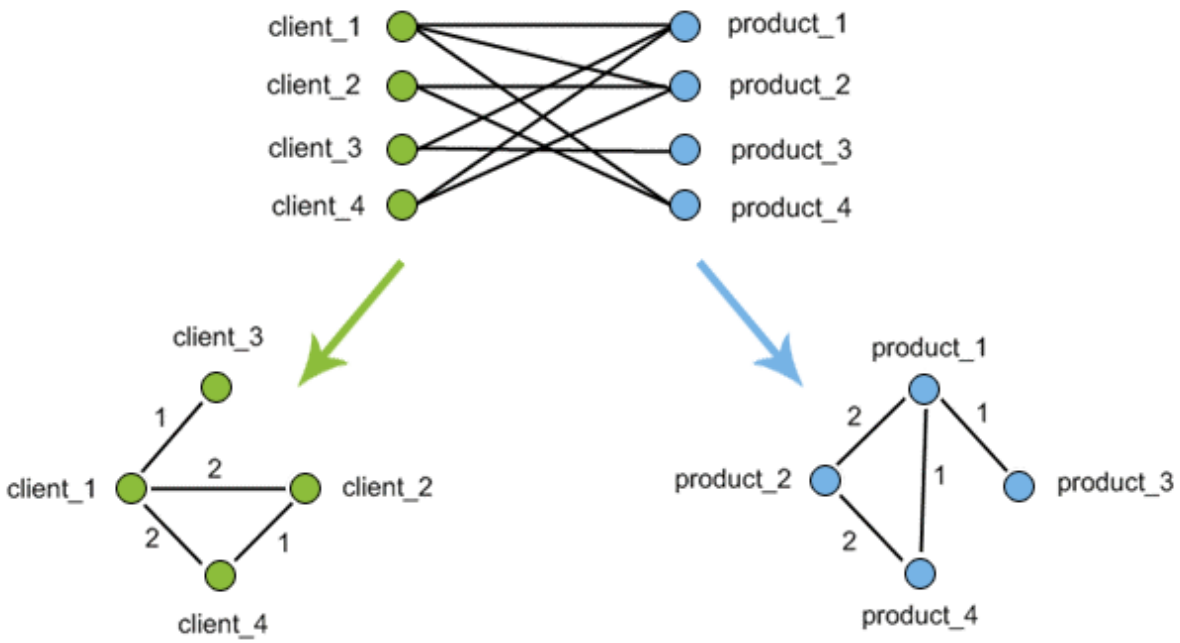
5.10.1.2.2 Transactions

Pour créer des graphes de type Transactions, le jeu de données des liens doit décrire les liens entre deux populations distinctes telles que "clients" et "produits".

CLIENT_ID	PRODUIT_ID	Date
1	1	03/12/2007
1	2	03/13/2007
1	4	03/14/2007
2	2	03/15/2007
2	4	03/16/2007
...	...	...

Le processus de chargement crée un graphe bipartite, dans lequel tous les liens se situent entre deux populations distinctes. Ce graphe permet de dériver deux graphes unipartites pour chaque population, tels que le graphe des clients ayant acheté n produits en commun et le graphe des produits ayant plus de n clients

en commun. Les liens de ces types de graphes peuvent être étiquetés selon le nombre de produits/clients en commun :



Exemples de graphes :

- Graphes liant les clients qui ont acheté au moins 3 produits en commun
- Graphes liant les produits ayant au moins 8 clients en commun

5.10.1.2.3 Plus proches voisins

Ce type de graphe est créé à partir d'un jeu de données contenant des entités et leur distance respective. Étant donné le paramètre k, l'objectif est d'obtenir un graphe reliant chaque nœud à ses plus proches voisins k :

Nœud 1	Nœud 2	Distance	Date
53123	63152	10	03/12/2007
68129	15487	12	03/13/2007
43261	32536	6	03/14/2007
12993	22866	2	03/15/2007
60765	63152	75	03/16/2007
18934	48765	23	03/17/2007

Il est possible d'utiliser la même condition de filtre que pour l'option Liens uniquement, à laquelle s'ajoute le paramètre k.

Remarque : ce type de graphe est toujours dirigé. L'explication réside dans l'exemple qui suit, où un graphe est créé et dans lequel chaque nœud est connecté à son propre voisin le plus proche :

A est le voisin le plus proche de B et est également le voisin le plus proche de C. Dans un graphe non dirigé, A serait lié à deux voisins différents. Cela revient à dire que A possède deux plus proches voisins, ce qui serait contradictoire. Cela est dû au fait que la relation "est le plus proche voisin de" n'est pas une relation symétrique.

Exemple de graphe :

- Graphe reliant chaque nœud à ses 3 plus proches voisins.

Informations associées

Liens uniquement [page 230]

5.10.1.2.4 Liens uniquement

Il s'agit ici du type de graphe le plus simple, dans lequel toutes les informations de liens sont déjà calculées et définies de manière explicite :

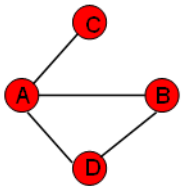
Nœud 1	Nœud 2	Type	Date
0143525691	0631527812	MESSAGE VOCAL	03/12/2007
0467892154	0154876556	MESSAGE VOCAL	03/13/2007
0499472604	0325363819	SMS	03/14/2007
0282812993	0228627344	SMS	03/15/2007
0186076517	0631527812	MMS	03/16/2007
0189340041	0154876556	MESSAGE VOCAL	03/17/2007

Il est possible de concevoir plusieurs graphes à la volée à partir du même jeu de données en créant différents filtres, tels que le filtrage par date, le filtrage de colonnes à l'aide de valeurs discrètes (sms, message vocal) ou de valeurs continues.

Exemples de graphes :

- Graphes de clients ne possédant que des messages vocaux.
- Graphe de clients possédant des communications par SMS entre le 03/01/2007 et le 04/01/2007

Node 1	Node 2
A	B
A	C
B	D
D	A



5.10.1.3 Personnalisation des ensembles de graphes

Les graphes décrits peuvent être aisément définis et plusieurs graphes peuvent être extraits, depuis le même jeu de données et en un seul balayage.

5.10.1.4 Dirigé / Non dirigé

Un graphe peut être non dirigé (représentant un lien symétrique : A est l'ami de B et B est l'ami de A) ou dirigé (représentant une relation non réflexive : A a appelé B ; le client A a acheté le produit C). Remarque : le nombre de liens peut varier entre un graphe dirigé et un graphe non dirigé.

5.10.1.5 Nœuds du réseau/hors réseau

Les nœuds hors réseau sont des nœuds du graphe pour lesquels aucune information complémentaire (âge, état civil, revenu) n'est disponible. Ils ne peuvent pas être liés à une base de données pour ajouter des attributs descriptifs, mais représentent tout de même un certain nombre d'informations. Dans le cas d'une compagnie de téléphone, par exemple, les nœuds hors réseau représentent les clients d'une autre compagnie et leur présence dans un voisinage de nœud peut aider à prédire le taux de résiliation.

À l'inverse, les nœuds du réseau sont ceux pour lesquels des informations complémentaires sont disponibles.

Les informations relatives aux nœuds du réseau et hors réseau se présentent en tant que méta-attributs pour les nœuds.

5.10.1.6 Détection de communautés

Ce mode de visualisation affiche les nœuds sous forme de membres de communautés. Chaque communauté est affichée avec une forme et une couleur différentes pour mieux les distinguer. Selon leurs connexions aux

autres nœuds d'une communauté identique ou distincte, les nœuds peuvent jouer des rôles différents au sein de leurs communautés. Les rôles en question sont les suivants :

- *Nœud passerelle* : nœud reliant des communautés entre elles en agissant comme un pont. Ce type de nœud comprend beaucoup de liens avec l'extérieur de sa communauté, mais peu à l'intérieur de celle-ci.
- *Nœud local* : nœud comprenant un grand nombre de liens dans une communauté, mais très peu de liens avec d'autres communautés.
- *Nœud social* : nœud comprenant un grand nombre de liens dans sa propre communauté, mais aussi avec d'autres communautés.
- *Nœud passif* : nœud comprenant un faible nombre de liens dans sa propre communauté et avec d'autres communautés.

5.10.1.6.1 Supercommunautés

L'application associe des petites sous-communautés à des intérêts ou comportements similaires pour créer de plus grandes "supercommunautés". Dans de nombreux réseaux sociaux, un individu ou un nœud ne peut être connecté qu'à un nombre limité d'autres nœuds (exemple : un numéro de téléphone est en moyenne connecté à environ 10 autres numéros de téléphone). Si vous êtes intéressé par la découverte de tendances ou l'implémentation d'un programme marketing viral de type réseau social, vous devez comprendre le concept dans son ensemble.

En détectant une supercommunauté, les utilisateurs métier peuvent explorer visuellement des réseaux sociaux, identifier rapidement des supercommunautés à influence sociale forte et utiliser des attributs sociaux (statistiques sur la communauté et la supercommunauté, indicateurs de centralité, etc.) pouvant renforcer les modèles prédictifs pour les initiatives de marketing viral.

5.10.1.7 Méga-hub

Un nœud peut être considéré comme un méga-hub lorsque son degré est supérieur à un certain seuil et qu'il possède donc une contribution négative selon la modularité globale du graphe et sa probabilité de mise en cluster. Un nœud connecté à un grand nombre de voisins tend à regrouper des communautés distinctes en une seule communauté principale, masquant ainsi les clusters sous-jacents actuels.

La suppression de ces nœuds hyper connectés peut améliorer de manière significative la durée de traitement et écarter les nœuds secondaires de l'analyse. Étant donné leur degré élevé, ces types de nœuds sont susceptibles de créer un "cluster artificiel" dans le réseau, rendant ainsi la détection des communautés difficile et masquant la structure réelle du graphe.

On rencontre ce type de nœuds méga-hub dans les réseaux de télécommunications (numéro de service, centre d'appels, taxis), les réseaux de transactions (produits phares, top des ventes). Ces nœuds possèdent de nombreux liens, mais n'ont, d'un point de vue sémantique, aucun intérêt.

La suppression des méga-hubs peut avoir des effets positifs :

- augmentation de la modularité obtenue
- réduction de la durée de traitement de la communauté (car l'attribution de communauté implique la recherche de voisins)

- réduction de la durée de détection d'appariement (car les voisins associés au méga-hub auront un cercle secondaire très large)
- réduction de la taille de la plus grande communauté (provoquée à tort par le fait que certains nœuds soient connectés à trop de nœuds.

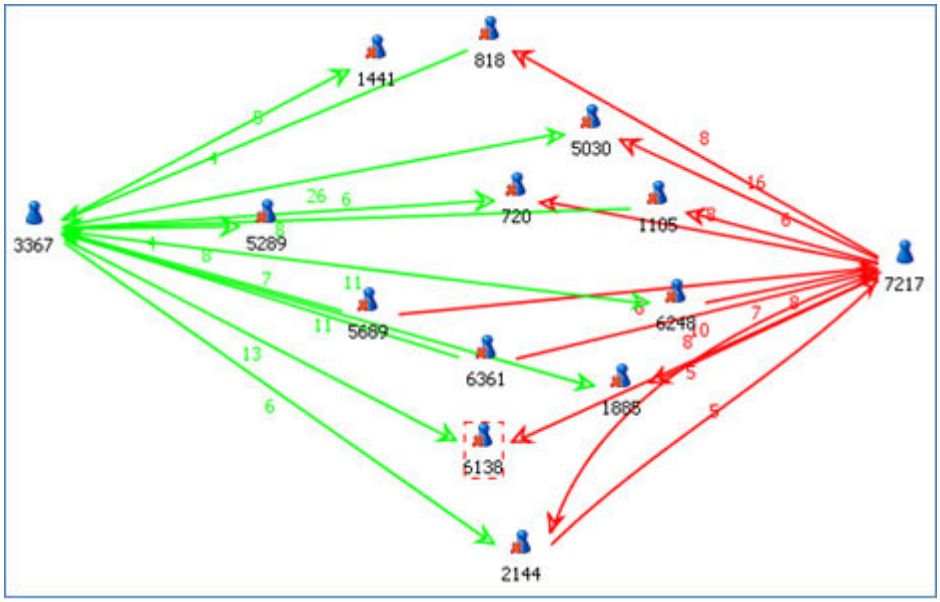
Informations associées

Détection de communautés [page 231]
Appariement des nœuds [page 233]

5.10.1.8 Appariement des nœuds

L'appariement des nœuds peut être utilisé pour identifier un élément (client, produit, etc.) présent dans deux graphes (avec deux nœuds distincts) grâce à leurs voisins communs. Pour une compagnie téléphonique, cette technique peut notamment être utilisée pour détecter le churn rotationnel (identifier les clients qui changent fréquemment d'abonnement téléphonique afin de bénéficier des offres de bienvenue).

Pour extraire un élément figurant dans deux graphes différents : les liens entre deux nœuds peuvent être étiquetés soit à l'aide du nombre de voisins communs (mode total), soit à l'aide du ratio common_neighbors/ average_neighbors_count (mode ratio). Cette valeur de ratio peut être présentée en tant que probabilité de correspondance : un ratio de 1 signifie que les deux nœuds considérés possèdent 100 % de voisins en commun et sont susceptibles de représenter un élément unique et identique.



Informations associées

Plus proches voisins [page 229]

5.10.1.9 Attributs dérivés

Une fois les graphes générés, Social peut les appliquer à un nouveau jeu de données afin de dériver la variable sociale de l'ensemble de graphes généré.

Mode voisinage

L'attribut dérivé est un nombre simple de voisins pour le nœud pris en compte. Il est également possible de connaître la somme/le ratio de voisins hors ligne dans le voisinage direct. Un graphe dirigé génère deux colonnes ; une pour le degré entrant et une pour le degré sortant.

Mode cercle

Le but est d'extraire les informations contenues dans le voisinage direct d'un nœud en agrégeant les propriétés de ses voisins directs. Il est alors possible de définir un profil en calculant la moyenne des variables continues (âge, revenu) ou le ratio/total de propriétés discrètes (état civil, sexe). Un graphe dirigé génère deux colonnes ; une pour le degré entrant et une pour le degré sortant.

Analyse de la centralité

Cet indicateur mesure le rôle et l'importance d'un nœud dans le graphe et son influence sur son voisinage.

Mode communauté

Le but est de détecter des communautés existantes : les attributs dérivés marquent les communautés potentielles et permettent d'identifier des rôles (comportement des nœuds à l'intérieur et à l'extérieur de leur communauté).

Appariement des nœuds

Les attributs générés peuvent être utilisés pour identifier des nœuds dans deux différents graphes qui représentent le même élément (client, produit, etc.). Ce mode est utile pour détecter une résiliation rotationnelle ou des utilisateurs multi-SIM par exemple.

Informations associées

Détection de communautés [page 231]
Nœuds du réseau/hors réseau [page 231]
Plus proches voisins [page 229]
Dirigé / Non dirigé [page 231]

5.10.2 Définition des paramètres

5.10.2.1 Quel type de graphe créer ?

Il existe trois manières de créer un graphe d'analyse des réseaux sociaux :

- *Construire un graphe social depuis un jeu de données*
  Cette option vous permet de sélectionner un jeu de données et de l'utiliser pour générer un nouveau graphe. Le jeu de données source peut être stocké dans un fichier ou une base de données.
- *Extraire un nouveau graphe depuis un graphe Social existant*
  Cette option vous permet de charger un modèle créé avec le module Social et d'en extraire un graphe défini dans ce modèle.
- *Projet un graphe à partir d'un graphe bipartite*
  Cette option vous permet de projeter un graphe à partir d'un graphe bipartite existant, c'est-à-dire de créer un graphe reliant les membres d'une même population en se basant sur les liens qu'ils ont avec l'autre population. Ainsi, dans un graphe bipartite reliant des clients à des produits, vous pouvez projeter un graphe reliant entre eux les produits qui ont été achetés par les mêmes clients.

> **i** Remarque
>
> Cette option n'est disponible que lorsqu'un graphe bipartite a déjà été créé dans le modèle en cours, soit à partir d'un jeu de données, soit par chargement à partir d'un modèle existant.

5.10.2.2 Définitions de graphe

-
-
-
-
-
- Paramétrage de la projection [page 239]
- Ajout de filtres de colonnes [page 242]
- Filtrer la population pour un graphe dérivé [page 243]

Lors de la création d'un graphe ou de la sélection d'un graphe dans la *Liste des graphes*, les paramètres du graphe s'affichent dans la partie droite du panneau. Ces paramètres dépendent du type de graphe sélectionné dans le panneau précédent.

Ce panneau vous permet de définir les informations dont vous avez besoin, par exemple :

- Un grand nombre de graphes identiques ayant des dates de début et de fin différentes vous permet de visualiser l'évolution d'une population.
- En télécommunication, il est possible de créer un graphe par type d'appel en mettant un filtre sur la colonne "Type d'appel".

Ajout de graphes et de filtres

1. Cliquez sur *Ajouter un graphe* pour créer un graphe dans la *Liste des graphes*. Un graphe par défaut est créé automatiquement.

   > **i** Remarque
   >
   > Lorsque vous extrayez un graphe d'un modèle social enregistré précédemment, les informations ci-dessous sont rappelées et ne peuvent pas être modifiées :
   >
   > - Le *Nom du graphe* et le *Type des liens* dans l'onglet *Paramètres du graphe*
   > - La *Source*, qui indique l'emplacement du modèle à partir duquel le graphe a été extrait, le nom du *Fichier/Table* contenant le modèle et le *Graphe source*, qui est le nom du graphe extrait dans l'onglet *Modèle*.
   >
   > Vous pouvez ajouter des filtres de colonne à un graphe extrait.

2. Cliquez sur *Dupliquer le graphe* et sélectionnez l'option souhaitée pour copier le graphe avec tous ses paramètres et filtres.
3. Cliquez sur *Ajouter un filtre de colonne* pour créer un filtre pour le graphe sélectionné.
4. Cliquez sur *Supprimer la sélection* pour supprimer l'élément sélectionné (graphe ou filtre).
5. Une fois tous les graphes et filtres définis, cliquez sur *Suivant*.

Définition de la configuration générale du graphe

La *Configuration générale* s'applique à tous les types de graphe.

1. Dans la *Liste des graphes*, sélectionnez le graphe pour lequel vous voulez modifier les paramètres. Lorsque vous créez un graphe, il est sélectionné par défaut.
2. Utilisez l'onglet *Jeu de données* pour afficher les informations relatives au jeu de données utilisé pour créer le graphe.
3. Dans l'onglet *Paramètres du graphe*, entrez le nom de votre graphe dans le champ *Nom du graphe*.
4. Sélectionnez le type de graphe que vous souhaitez créer dans le *Mode de création des graphes*. Le mode de création des graphes doit être défini en fonction de votre jeu de données :

Cette option...	génère des graphes...	contenant les informations suivantes sur le lien...
Liens uniquement	qui lient les identifiants de la colonne source avec ceux de la colonne cible.	aucun
Contacts	qui contiennent uniquement les liens trouvés au moins un certain nombre de fois.	nombre d'occurrences du lien
Plus proches voisins	qui comportent un nombre spécifique de voisins les plus proches pour chaque nœud. Les plus proches voisins sont déterminés par leur distance relative qui doit être renseignée dans une colonne supplémentaire.	distance
Transactions	qui comportent les relations entre des objets de nature différente, comme par exemple des clients et des produits.	
Proximité	Graphe servant à la géolocalisation	

Définition des rôles de colonnes

1. Dans la liste *Nœud source*, sélectionnez la colonne contenant l'identifiant du nœud source lors de la création d'un graphe dirigé ou la première population lors de la création d'un graphe bipartite.
2. Dans la liste *Nœud cible*, sélectionnez la colonne contenant l'identifiant du nœud cible lors de la création d'un graphe dirigé ou la seconde population lors de la création d'un graphe bipartite.

> **i** Remarque
>
> Le *Nœud source* et le *Nœud cible* doivent être différents et leur ordre détermine la manière dont les liens s'affichent lorsque vous visualisez le graphe.

> **i** Remarque
>
> Pour utiliser les fonctionnalités de géolocalisation, vous devez définir une variable de position que vous pouvez créer à partir des variables de latitude et de longitude présentes dans votre jeu de données.

> Vous pouvez créer la variable de position à l'aide de l'option *Variables composites* dans le panneau *Description du jeu de données des liens*.

3. Facultatif : Si besoin, sélectionnez la colonne contenant les informations de date dans la liste *Colonne de date*. Cette colonne peut être utilisée pour filtrer les données à utiliser dans le graphe.
4. Pour le type de graphe *Contact* et *Liens uniquement*, dans la liste *Type des liens*, choisissez si la direction des liens est pertinente pour votre problématique. Sélectionnez *Dirigé* si vous souhaitez que seuls les liens allant des nœuds source aux nœuds cible soient pris en compte et *Non dirigé* dans tous les autres cas. Par conséquent, il est possible qu'il y ait une différence entre le nombre d'événements d'un graphe dirigé et d'un graphe non dirigé.
5. Pour le type de graphe *Voisins*, sélectionnez la distance relative entre les nœuds dans la liste *Colonne de distance*. Cela peut être un temps de communication, un nombre d'articles achetés, une distance physique, etc.
6. Pour les types de graphe *Transactions* et *Proximité*, vous pouvez filtrer vos résultats en fonction de la valeur d'une variable spécifique dans votre jeu de données.
   Par exemple, si vous créez des règles permettant de suggérer des produits complémentaires aux clients, vous pouvez choisir de ne suggérer que des produits qui sont en stock.
   1. Sélectionnez la variable que vous souhaitez utiliser pour filtrer le graphe de population dans la liste *Cible*.
   2. Saisissez la valeur que vous voulez utiliser pour filtrer les résultats dans le champ *Clé cible*. Si la variable que vous avez sélectionnée comme *Cible* peut avoir plusieurs valeurs, seules les règles pour lesquelles la valeur est celle indiquée dans le champ *Clé cible* seront fournies.

Définition des conditions de création des liens

Selon le type de graphe que vous créez, les conditions de création des liens entre les nœuds varient. Dans la section *Conditions de création des liens*, définissez les options appropriées comme indiqué dans le tableau suivant.

Paramètre	Utilisez ce paramètre pour...	Type de graphe
Nombre minimum de contacts	Définir le nombre minimal de fois où doit apparaitre le couple nœud source-nœud cible dans le jeu de données pour pouvoir créer un lien.	• Contacts
Utiliser une colonne de poids	Affecter un poids différent à chaque lien. Dans le cas d'un graphe Contacts, l'association d'une colonne de poids et d'un poids minimal peut être utilisée pour déterminer quels liens doivent être créés.	• Contacts • Liens uniquement • Transactions
Poids minimal	Définir le poids minimal en dessous duquel le lien ne sera pas créé	• Contacts

Paramètre	Utilisez ce paramètre pour...	Type de graphe
Nombre maximal de connexions	Indiquer le nombre maximal de connexions créées pour chaque nœud. Lorsque le nombre de maximal de connexions est atteint pour un nœud, plus aucune connexion n'est créée pour ce nœud. Si le seuil pour la détection des méga-hubs est inférieur à la valeur par défaut de ce champ (50 0000), saisissez-le ici ainsi que dans les paramètres de la détection des méga-hubs dans le panneau suivant.	• Contacts • Liens uniquement • Plus proches voisins • Transactions
Nombre maximal de voisins	Définir le nombre maximal de plus proches voisins à conserver pour chaque nœud	• Plus proches voisins

Définition de la période

Utilisez cette section pour filtrer les événements sur une période de temps déterminée. Vous pouvez définir une date de début, une date de fin ou les deux.

> **i** Remarque
>
> La date de début est incluse dans la période définie, et la date de fin en est exclue.

1. Si vous voulez utiliser uniquement les événements qui se produisent après une date spécifique :
   1. Cochez *Date de début*.
   2. Cliquez sur le bouton calendrier situé à droite du champ de date pour sélectionner la date.
2. Si vous voulez utiliser uniquement les événements qui se produisent avant une date spécifique :
   1. Cochez *Date de fin*.
   2. Cliquez sur le bouton calendrier situé à droite du champ de date pour sélectionner la date.

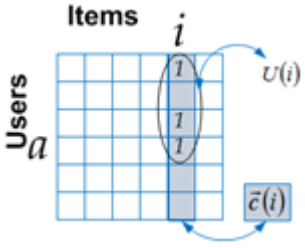
5.10.2.2.1 Paramétrage de la projection

Lorsque vous dérivez un graphe à partir d'un graphe bipartite vous devez définir une projection de ce graphe pour une de ses entités (colonne source ou cible).

Le tableau suivant liste les paramètres disponibles pour la projection du graphe :

Paramètres	Description
Graphe à projeter	Nom du graphe bipartite d'où est issue la projection

Paramètres	Description
Entité	Entité utilisée comme nœud pour le graphe en cours. Un graphe bipartite crée des liens entre deux types d'entités, une seule pouvant être utilisée dans un graphe classique.
Garder les N les plus importants	Le nombre d'appariement que vous souhaitez garder parmi les mieux classés.

Paramètres	Description
Poids	Indique quelle valeur utiliser comme poids pour les liens. Il est recommandé d'utiliser l'*indice de Jaccard* ou le *Ratio d'indépendance*, ce dernier en particulier lorsque vous travaillez avec des communautés. *Support* : nombre de liens trouvés pour chaque nœud. *Indice de Jaccard* : Permet d'évaluer la similarité entre des ensembles. Il est défini comme la taille de l'intersection divisée par la taille de l'union des ensembles : $$J(A,B) = \frac{\lvert A \cap B \rvert}{\lvert A \cup B \rvert}$$ *Ratio d'indépendance* : Deux événements A et B sont indépendants si la probabilité que les deux événements se produisent est égale à la probabilité de l'événement A fois la probabilité de l'événement B : $$\Pr(A \cap B) = \Pr(A).\Pr(B)$$ Le poids est alors calculé suivant la formule ci-dessous : $$w = \frac{\Pr(A \cap B)}{\Pr(A).\Pr(B)}$$ Un poids de 1 indique deux événements complètement indépendants. Plus le poids est élevé, plus les événements sont corrélés. *Cosinus*  où : a : l'utilisateur actif i et j : deux éléments $U(i)$: le nombre d'utilisateurs achetant i $\vec{c}(i)$: la colonne de la matrice R pour l'élément i $$Sim(i,j) = cos[\vec{c}(i),\vec{c}(j)] = \frac{\vec{c}(i) \bullet \vec{c}(j)}{\lVert \vec{c}(i) \rVert_2 \times \lVert \vec{c}(j) \rVert_2} = \frac{\sum_{u=1}^{L} w_{ui} \times w_{uj}}{\sqrt{\sum_{u=1}^{L} (w_{ui})^2} \times \sqrt{\sum_{u=1}^{L} (w_{uj})^2}}$$

Paramètres	Description
Nombre maximum d'itérations	Nombre maximum d'itérations avant que l'algorithme s'arrête si les critères Epsilon ne sont pas atteints.
Privilégier la consommation de la mémoire	L'utilisation de cette option réduit l'utilisation de la mémoire. Notez que cela ralentit quelque peu le processus.
Privilégier la vitesse de traitement	Utilisez cette option pour accélérer le processus. Notez que cela augmentera l'utilisation de la mémoire.
Support minimal	Le support est le nombre d'éléments que deux entités ont en commun. Ainsi, lorsque vous projetez un graphe bipartite reliant des produits et des clients, le support du lien joignant le produit A et le produit B est le nombre de clients ayant acheté les deux produits. Un lien dont le support est inférieur au *Support minimal* ne sera pas créé.

5.10.2.2.2 Ajout de filtres de colonnes

Ajouter un filtre

1. Cliquez sur le bouton *Ajouter un filtre de colonne*. Les *Paramètres de filtrage* sont affichés dans la partie droite du panneau.
2. Sélectionnez la colonne sur laquelle sera appliqué le filtre dans la liste *Colonne filtrée*.
3. Si la colonne sélectionnée contient un nombre ou une indication de temps, vous pouvez définir une plage de valeurs comme filtre. Dans ce cas :
   a. Cochez la case d'option *Accepter les valeurs dans l'intervalle*.
   b. Indiquez la valeur minimale pour l'intervalle de votre filtre dans le champ *Minimum*. La valeur minimale est incluse dans l'intervalle.
   c. Indiquez la valeur maximale pour l'intervalle de votre filtre dans le champ *Maximum*. La valeur maximale est exclue de l'intervalle.
4. Vous pouvez aussi saisir une ou plusieurs valeurs à accepter ou à rejeter :
   a. Cochez la case d'option *Accepter* ou *Rejeter.*
   b. Cliquez sur + pour ajouter une valeur.
   c. Entrez la valeur ou utilisez les flèches.
   d. Cliquez sur - pour enlever la valeur sélectionnée de la liste.

5.10.2.2.3 Filtrer la population pour un graphe dérivé

Pour les graphes dérivés, vous pouvez utilisez les options fournies dans l'onglet *Liste des candidats* afin de filtrer la population que vous souhaitez afficher dans le graphe.

- Pour utiliser l'*Événement cible requis*, vous devez premièrement indiquer une variable *Cible* et une valeur de *Clé cible* dans les paramètres du graphe de transactions que vous souhaitez utiliser pour créer le graphe dérivé (voir la section Rôles de colonnes [page 237] ci-dessus). Si la variable *Cible* n'est pas définie au préalable, aucune règle ne peut être générée.
- L'option *Méthode de sélection* et l'option *Limite de transition suivante* sont disponibles uniquement si le graphe dérivé a été créé en se basant sur un graphe de proximité à l'aide d'une colonne de poids.

1. Cochez l'option *Événement cible requis* pour filtrer la population du graphe actuel sur la valeur de la variable cible définie pour le graphe bipartite.
2. Pour définir l'option *Méthode de sélection* :
   a. Sélectionnez comment filtrer la population en fonction de la période dans la liste *Méthode de sélection*.

Utilisez la méthode de sélection...	Pour...
Tout	sélectionner toute la population sans heure définie
Synchronisme	sélectionner les personnes qui étaient au même endroit (même tuile) au même moment
Séquence	sélectionner les personnes qui ont suivi le même chemin. Notez que pour cette méthode de sélection, l'*Intervalle de temps maximum* correspond au décalage entre le début et la fin du chemin.

   b. Si vous avez sélectionné la méthode de sélection *Synchronisme* ou *Séquence*, cliquez sur *Échelle de temps* pour sélectionner l'unité de temps à utiliser.

3. Entrez le nombre maximum de transitions à prendre en compte après avoir trouvé une paire dans le champ *Limite de transition suivante*.

5.10.2.3 Post-traitement

Ce panneau vous permet de sélectionner le type de post-traitement que vous souhaitez appliquer sur les graphes que vous souhaitez créer.

- Post-traitement générique [page 244]
- Appariement des nœuds [page 246]

5.10.2.3.1 Post-traitement générique

Détection de communautés

Détection de communautés

Cette option permet d'identifier les communautés existant dans le jeu de données. Une communauté peut être considérée comme un sous-graphe dont la densité des connexions internes est supérieure à celle de ses connexions vers le reste des nœuds du réseau. La détection des communautés utilise un algorithme itératif dont les critères d'arrêt sont le **nombre maximum d'itérations** et le **gain de modularité d'une itération sur l'autre**, représenté par Epsilon.

Pour en savoir plus sur le calcul de la détection de communautés, voir les informations sur l'algorithme de Louvain .

Vous pouvez également **baser les communautés d'un graphe sur celles d'un autre graphe**. Ainsi, les communautés du mois de février peuvent être initialisées sur celle de janvier, vous permettant de visualiser leur évolution d'un mois sur l'autre.

- Pour activer la détection des communautés
  - Cochez la case **Détection des communautés** correspondant au graphe pour lequel vous voulez activer la détection.
    Remarque : La détection des communautés n'est pas disponible pour les graphes transactionnels.
- Pour configurer la détection des communautés :
  1. Cliquez sur le bouton **...** situé à droite de la case à cocher **Détection des communautés**. La fenêtre **Configuration de la détection des communautés** s'affiche.
  2. Cochez la case **Activer la détection des communautés**.
  3. Définissez les options que vous souhaitez appliquer. Le tableau suivant résume les options disponibles :

Option	Description
Nombre maximum d'itérations	Nombre maximum d'itérations avant que l'algorithme ne s'arrête si les critères Epsilon ne sont pas atteints.
Epsilon	Valeur en-dessous de laquelle le gain de modularité n'est plus suffisamment intéressant pour justifier une autre itération de l'algorithme.
Se baser sur les communautés de	graphe à partir duquel vous pouvez utiliser les communautés comme base pour le graphe en cours. La liste Se baser sur les communautés de comprend uniquement les graphes pour lesquels la détection des communautés a été activée. La base de communautés n'est pas disponible pour les graphes transactionnels.

Option	Description
Ignorer les auto-liens	Réduit le nombre de communautés et apporte une meilleure perspective de la structure du graphe.
Désactiver les supercommunautés	Désactive la création de supercommunautés, ce qui accélère la création lors de l'utilisation de grands graphes.

4. Cliquez sur **OK** pour valider les paramètres.

Optimisation de la détection des méga-hubs

L'option *Optimiser la détection des méga-hubs* supprimera tous les nœuds dont le degré de connectivité est supérieur au seuil défini. Il existe deux façons de définir ce seuil :

- Utiliser le seuil automatique, basé sur la distribution des nœuds en fonction de leur degré de connectivité. Le seuil automatique est calculé à l'aide de la formule suivante : moyenne + facteur de l'écart-type * écart-type

  > **i** Remarque
  >
  > Par défaut, le facteur de l'écart-type est fixé à 4.

- Utiliser une valeur fixe.
- Pour activer l'optimisation de la détection des méga-hubs
  - Cochez la case *Optimiser la détection des méga-hubs*.
- Pour paramétrer la détection des méga-hubs

> **i** Remarque
>
> Lorsque vous paramétrez cette option pour un graphe transactionnel, vous devez configurer le seuil pour les deux populations.

1. Cliquez sur le bouton ... situé à droite de la case à cocher *Optimisez la détection des méga-hubs*. La fenêtre *Configuration de la détection des méga-hubs* s'affiche.
2. Cochez la case *Optimiser la détection des méga-hubs*.
3. Sélectionnez le type de seuil que vous souhaitez utiliser :
   - *Seuil automatique*
   - *Seuil manuel*
4. Selon le type de seuil sélectionné :
   - pour le *seuil automatique*, entrez une valeur dans le champ *Facteur de l'écart-type* pour paramétrer la formule utilisée lors du calcul ;
   - utilisez le champ *Seuil d'exclusion* pour définir le *seuil manuel*.
5. Cliquez sur *OK* pour valider les paramètres.

5.10.2.3.2 Appariement des nœuds

L'appariement des nœuds peut être utilisé pour identifier un élément (client, produit, etc.) présent dans deux graphes (avec deux nœuds distincts) grâce à leurs voisins communs. Pour une compagnie téléphonique, cette technique peut notamment être utilisée pour détecter le churn rotationnel (identifier les clients qui changent fréquemment d'abonnement téléphonique afin de bénéficier des offres de bienvenue).

- Pour apparier les graphes :
  1. Sélectionnez l'onglet *Appariement*.
  2. Cliquez sur le bouton situé en haut à droite de l'onglet pour ajouter une nouvelle définition d'appariement. La fenêtre *Définition de l'appariement* s'ouvre.
  3. Sélectionnez le *Premier graphe*.
  4. Sélectionnez le *Second graphe*.
  5. Cochez l'option située à côté du champ *Nom du graphe en sortie* et saisissez un nom si vous souhaitez que les graphes appariés s'affichent comme un nouveau graphe.
  6. Dans la section *Conditions d'appariement*, saisissez un *nombre minimal de voisins communs*.
  7. Cochez l'option située à côté du champ *Proportion minimale de voisins communs* et indiquez une valeur si vous souhaitez définir cette condition.
  8. Cliquez sur *OK* pour valider la définition d'appariement. La fenêtre se ferme et le nouvel appariement s'affiche dans la liste.
  9. Si vous souhaitez définir un autre appariement, refaites les étapes 2 à 8.

Conditions d'appariement

Condition	Description
Nombre minimal de voisins communs	Deux nœuds sont appariés uniquement s'ils ont un nombre de voisins communs égal ou supérieur au nombre indiqué.
Garder les N plus importants	Seules les N paires les plus importantes sont conservées. L'importance des paires est déterminée par le *Type d'appariement*. Lorsque la valeur de ce champ est égale à 0, aucun filtre n'est appliqué. Les graphes conservent les N liens sortants ayant les poids les plus importants, donc lorsque l'utilisateur veut générer les graphes contenant la proportion et le nombre de voisins communs, il se peut que des liens vers des paires différentes soient conservés.
Proportion minimale de voisins communs	Seules les paires pour lesquelles l'indice de Jaccard est supérieur ou égal au nombre indiqué sont conservées.

Type d'appariement	• Ratio • Jaccard • Ratio d'indépendance (uniquement dans le cas d'appariements dans un même graphe) • Confiance • Coefficient de segmentation Les formules correspondant à chacun de ces types sont indiquées dans la section ci-dessous.
Proportion pondérée	La métrique utilisée pour calculer le *Type d'appariement* est pondérée avec la variable de poids déclarée lors de la définition du graphe.
Inclure le graphe du "Nombre de voisins communs"	Lors de la génération de graphe, un graphe supplémentaire est créé dans lequel les liens sont étiquetés avec le nombre de voisins communs.

Types d'appariement

- **Ratio**

Non-pondéré	$$ratio(A,B) = \frac{\#\,common_neighbors(A,B)}{(\#\,neighbors(A)+\#\,neighbors(B))/2}$$
Pondéré	$$ratio(A,B) = \frac{w_degree_common(A,B)}{(w_degree(A) + w_degree(B))}$$

- **Jaccard**

Non-pondéré	$$jaccard(A,B) = \frac{\#\,common_neighbors(A,B)}{\#\,neighbors(A)+\#\,neighbors(B)-\#\,common_neighbors(A,B)}$$
Pondéré	$$jaccard(A,B) = \frac{w_degree_common(A,B)}{w_degree(A) + w_degree(B)}$$

- **Ratio d'indépendance**

Non-pondéré	$$independence_ratio(A,B) = \frac{P(A\,\&\,B)}{P(A)P(B)} = \frac{\#\,common_neighbors(A,B)*\#\,links}{\#\,neighbors(A)*\#\,neighbors(B)}$$
Pondéré	$$independence_ratio(A,B) = \frac{P(A\,\&\,B)}{P(A)P(B)} = \frac{w_degree_common(A,B) * total_weight}{w_degree(A) * w_degree(B)}$$

- **Confiance**

Non-pondéré	$$confidence(A,B) = \frac{P(A\,\&\,B)}{P(A)} = \frac{\#\,common_neighbors(A,B)}{\#\,neighbors(A)}$$

Pondéré	$confidence(A,B) = \frac{P(A \& B)}{P(A)} = \frac{w_degree_common_from(A)to(A,B)}{w_degree(A)}$

- **Coefficient de segmentation**

Non-pondéré	$clustering(A,B) = \frac{2 * (\#\ triangles(A) + \#\ common_neighbors(A,B))}{\#\ neighbors(A) * (1 + \#\ neighbors(A))}$
Pondéré	$clustering(A,B) = \frac{2 * (\#\ triangles(A) + w_degree_common_from(A)to(A,B))}{w_degree(A) * (1 + w_degree(A))}$

5.10.2.4 Conversion des identifiants

Le *jeu de données de conversion des identifiants* peut être utilisé pour relier le jeu de données des attributs descriptifs au *jeu de données de liens*. Si vous n'avez pas de jeu de données d'attributs descriptifs, vous pouvez utiliser ce type de jeu de données pour indiquer s'il existe des informations supplémentaires sur le nœud. *Ce jeu de données doit comporter deux colonnes d'identifiants*.

- Pour ajouter un jeu de données de conversion des identifiants
  1. Si vous avez défini un graphe *Transaction*, sélectionnez l'onglet correspondant au jeu de données pour lequel vous souhaitez ajouter un jeu de données de conversion des identifiants.
  2. Cochez la case *Ajouter un jeu de données de conversion des identifiants*.
  3. Cliquez sur le bouton *Sélectionner un jeu de données*. Une fenêtre s'ouvre vous permettant de choisir le jeu de données.
  4. Dans la liste *Type de données*, sélectionnez le type de la source de données.
  5. Cliquez sur le bouton *Parcourir* correspondant au champ *Dossier* pour choisir l'emplacement du jeu de données (il peut s'agir d'un dossier ou d'une base de données).
  6. Cliquez sur le bouton *Parcourir* correspondant au champ *Jeu de données* pour choisir le jeu de données (il peut s'agir d'un fichier ou d'une table de base de données).
  7. Les boutons *Voir les données* ou vous permettent de visualiser les données et de calculer les statistiques du jeu de données sélectionné.
  8. Le bouton *Paramètres avancés* vous permet de définir certains paramètres avancés tels que l'échantillonnage ou les paramètres de fichiers.
  9. Pour décrire les données vous pouvez :
     - soit laisser l'application deviner la description en cliquant sur le bouton *Analyser*,
     - soit utiliser une description existante en cliquant sur le bouton *Ouvrir*.
  10. Cliquez sur *OK*. Le nom du jeu de données sélectionné s'affiche dans le champ grisé à côté du bouton **Sélectionner le jeu de données**.
  11. Dans la liste *Identifiant original*, sélectionnez la colonne qui contient l'identifiant de nœuds provenant du jeu de données de liens.
  12. Dans la liste *Identifiant converti*, sélectionnez la colonne qui contient l'identifiant correspondant à celui qui existe dans le jeu de données d'enrichissement du nœud.

> **i** Remarque
>
> Les deux identifiants doivent avoir le même format de stockage.

5.10.2.5 Attributs descriptifs

Le **jeu de données d'attributs descriptifs** peut être utilisé pour donner des informations supplémentaires sur les variables du *jeu de données de liens*. S'il n'a pas la même colonne d'identifiants que le *jeu de données de liens*, vous devez sélectionner également un *jeu de données de conversion des identifiants*.

- Pour ajouter un jeu de données d'attributs descriptifs
  1. Sélectionnez l'onglet correspondant au jeu de données pour lequel vous souhaitez ajouter des attributs descriptifs.
  2. Cochez la case *Ajouter un jeu de données d'attributs descriptifs*.
  3. Cliquez sur le bouton *Sélectionner un jeu de données*. Une fenêtre s'ouvre vous permettant de choisir le jeu de données.
  4. Dans la liste déroulante, sélectionnez le *type de données*.
  5. Cliquez sur le bouton *Parcourir* correspondant au champ *Dossier* pour choisir l'emplacement du jeu de données (il peut s'agir d'un dossier ou d'une base de données).
  6. Cliquez sur le bouton *Parcourir* correspondant au champ *Jeu de données* pour choisir le jeu de données (il peut s'agir d'un fichier ou d'une table de base de données).
  7. Les boutons *Voir les données* ou vous permettent de visualiser les données et de calculer les statistiques du jeu de données sélectionné.
  8. Le bouton *Paramètres avancés* vous permet de définir certains paramètres avancés tels que l'échantillonnage ou les paramètres de fichiers.
  9. Pour décrire les données vous pouvez :
     - soit laisser l'application deviner la description en cliquant sur le bouton *Analyser*,
     - soit utiliser une description existante en cliquant sur le bouton *Ouvrir*.
  10. Cliquez sur *OK*. Le nom du jeu de données sélectionné s'affiche dans le champ grisé à côté du bouton *Sélectionner le jeu de données*.
  11. Dans la liste déroulante *Identifiant de nœud*, sélectionnez la colonne qui contient l'identifiant à utiliser pour relier les attributs descriptifs au jeu de données principal.

      > **i** Remarque
      >
      > Les identifiants des deux jeux de données doivent avoir le même format de stockage.

  12. Dans la liste *Colonnes disponibles*, sélectionnez les colonnes que vous voulez utiliser dans le graphe en cliquant sur le bouton >. Les attributs sélectionnés apparaissent alors dans la liste *Attributs descriptifs sélectionnés*.
  13. Vous pouvez ordonner la liste des attributs descriptifs en utilisant les boutons et ou rechercher des attributs par nom en utilisant le bouton et en saisissant tout ou partie du nom dans le champ qui s'affiche.
  14. Pour retirer un nom de colonne de la liste, sélectionnez-le et cliquez sur le bouton **<**.

5.10.2.6 Récapitulatif des paramètres de modélisation

Cet écran résume les paramètres de modélisation. Vous pouvez spécifier les paramètres suivants :

- le champ *Nom du modèle* est rempli automatiquement. Il correspond au nom de la variable cible, suivi d'un trait de soulignement ("_") et du nom de la source de données, sans son extension.
- la *Description* qui, une fois le modèle enregistré, permet de le reconnaître quand vous voulez le rouvrir avec SAP Predictive Analytics.

5.10.3 Utilisation du modèle

5.10.3.1 Options autonomes Social

Cet écran affiche les différentes options proposées pour visualiser, appliquer, enregistrer ou exporter le modèle.

Affichage

- *Aperçu du modèle* : affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.
- *Tables d'analyse* : ensembles de tableaux vous permettant une analyse détaillée de votre modèle.
- *Visualisation des nœuds* : affiche le graphe pour un nœud spécifique.
- *Recommandations* : vous permet de générer des recommandations à partir d'un graphe projeté d'un graphe bipartite.
- *Co-localisations* : lorsque vous utilisez des données contenant des informations géographiques, affiche la liste des couples d'éléments se trouvant au même endroit au même moment.
- *Visualiser les séquences* : lorsque vous utilisez des données contenant des informations géographiques, vous permet d'identifier les chemins et de les visualiser dans Google Earth.

Exécution

- *Application d'un modèle* : un nouveau jeu de données peut être scoré en utilisant le modèle créé. Les scores peuvent également être appliqués aux données utilisées pour créer le modèle.
- *Enregistrer le modèle* : le modèle créé peut être enregistré et rechargé ultérieurement.
- *Exporter le script KxShell* : génère un script KxShell qui peut être utilisé pour recréer le modèle.

- *Recommandations en masse* Génère une liste de produits recommandés pour chaque client de votre jeu de données.

5.10.3.2 Mode Application in-database

Ce mode optimisé permet d'appliquer le modèle directement dans la base de données. Cela évite d'avoir à extraire les données de la base de données et accélère le processus d'écriture des données de sortie du modèle.

Le mode Application in-database est disponible pour :

- Modèles de classification/régression avec un ordre du polynôme égal à 1.
- Modèles de segmentation générés avec des expressions SQL
- Recommandation et modèles sociaux

Vous pouvez utiliser ce mode si :

- Le jeu de données d'application d'entrée (table, vue, requête, manipulation de données) et le jeu de données de résultats (tables) provient de la même base de données.
- Au moins une variable du jeu de données d'entraînement utilisé pour créer le modèle est déclarée comme clé dans Automated Analytics.

  > **i** Remarque
  >
  > La (Les) clé (s) définie(s) dans le modèle est (sont) identifiée(s) en tant que clé primaire dans la table de sortie.

- L'utilisateur Automated Analytics qui déclenche la tâche d'application dispose des droits de lecture de l'entrée d'application et d'écriture de sortie d'application dans la base de données.
- Il existe un générateur d'indicateurs pour la base de données.
- Recommandation et modèles sociaux uniquement : Le modèle doit être enregistré dans la base de données contenant les jeux de données d'application.

  > **i** Remarque
  >
  > Si le modèle n'a pas été enregistré, une boîte de dialogue s'affiche et vous propose de l'enregistrer. Si vous n'enregistrez pas le modèle, Automated Analytics passe automatiquement au processus d'application standard.

> **i** Remarque
>
> Lorsque vous cochez l'option *Utiliser l'application directe dans la base de données*, l'option *Ajouter un écart de score* est automatiquement sélectionnée. Vous pouvez ainsi vérifier, pour chaque variable et chaque catégorie de variable, les écarts entre le modèle et le jeu de données d'entrée utilisé pour l'application du modèle.

5.10.3.3 Paramètres d'affichage

- Format des nœuds [page 252]
- Étiquettes [page 253]
- Info-bulles [page 254]

5.10.3.3.1 Format des nœuds

Il existe deux manières de définir le format des nœuds sur les graphes :

- Le *format simple* vous permet d'utiliser la taille des nœuds pour visualiser les valeurs d'une variable continue et la couleur des nœuds pour visualiser les valeurs d'une variable nominale ou continue.
- Le *format conditionnel* vous permet d'appliquer un format spécifique à chaque nœud selon un ensemble de règles que vous définissez.

Format simple

- Pour définir un format simple :
  - Cochez le bouton radio *Format simple*.
- Pour définir la taille des nœuds :
  1. Dans la liste *Taille en fonction de*, sélectionnez la variable que doit illustrer la taille des nœuds. Si vous ne souhaitez pas utiliser cette fonctionnalité, sélectionnez l'option *Rien*.

     > **i** Remarque
     >
     > Seules les variables continues sont proposées dans cette liste.

  2. Dans le champ *Taille minimum*, saisissez le diamètre minimal d'affichage des nœuds. Cette taille est définie en nombre de pixels et peut être comprise entre 5 et 150.
  3. Dans le champ *Taille maximum*, saisissez le diamètre maximal d'affichage des nœuds. Cette taille est définie en nombre de pixels et peut être comprise entre 5 et 150.
- Pour définir la couleur des nœuds :
  1. Dans la liste *Couleur en fonction de*, sélectionnez la variable que doit illustrer la couleur des nœuds. Si vous ne souhaitez pas utiliser cette fonctionnalité, sélectionnez l'option *Rien*.

     > **i** Remarque
     >
     > Seules les variables nominales et continues sont proposées dans cette liste.

  2. Si vous sélectionnez une *variable continue* :
     1. Cliquez sur le bouton situé à droite de l'intitulé *Couleur pour la valeur minimum*.
     2. Sélectionnez la couleur qui illustrera la valeur minimum de la variable.
     3. Cliquez sur le bouton situé à droite de l'intitulé *Couleur pour la valeur maximum*.

4. Sélectionnez la couleur qui illustrera la valeur maximum de la variable.
   La couleur de chaque nœud sera tirée du dégradé allant de la couleur sélectionnée pour la valeur minimum à la couleur sélectionnée pour la valeur maximum selon la valeur de la variable pour ce nœud.

3. Si vous sélectionnez une *variable nominale* :

   1. Cliquez sur le bouton + pour ajouter une valeur ou bien sur le bouton pour extraire les catégories de variables du jeu de données. Une couleur est associée à chaque catégorie de la variable sélectionnée.
   2. Vous pouvez modifier la couleur de chaque catégorie en cliquant sur la couleur associée.

- Pour modifier un format simple :
  - Cliquez deux fois sur le format que vous souhaitez modifier dans la liste *Légende des graphes*.

Format conditionnel

- Pour définir un format conditionnel :
  1. Cochez la case d'option *Format conditionnel*.
  2. Cliquez **+** pour ajouter un nouveau format conditionnel. Entrez le nom du format conditionnel, qui sera affiché dans la *Légende des graphes*.
  3. Entrez le nom du nouveau format, qui apparaîtra dans la *Légende des graphes*.
  4. Sélectionnez le mode d'affichage du nœud : forme colorée ou icône.
  5. Cliquez sur + pour ajouter les conditions qui définiront la mise en forme à appliquer aux nœuds.
  6. Utilisez les cases d'option situés sous la liste *Conditions* pour indiquer quand appliquer le format.
  7. Cliquez sur *OK* pour appliquer ces paramètres. La fenêtre *Définir un format conditionnel* se ferme.
  8. Répétez les étapes 2 à 7 pour chaque format conditionnel à définir.
  9. Cliquez *OK* pour valider.
- Pour modifier le format conditionnel :
  - Cliquez deux fois sur le format que vous souhaitez modifier dans la section *Légende des graphes* située à gauche du panneau *Visualisation des nœuds*.

5.10.3.3.2 Étiquettes

Vous pouvez choisir les informations à afficher sur le graphe pour chaque nœud.

- Pour modifier les étiquettes affichées sur le graphe :
  1. Dans la liste située à gauche, sélectionnez l'option *Étiquettes*.
  2. Dans la liste *Attributs disponibles*, sélectionnez les informations que vous souhaitez afficher sur le graphe.
     - Pour sélectionner et ajouter plusieurs attributs simultanément :
       1. Maintenez la touche Ctrl enfoncée.
       2. Sélectionnez les attributs à ajouter.
       3. Cliquez sur le bouton > pour ajouter les attributs à la liste *Attributs sélectionnés*.

- Pour trouver et sélectionner des attributs par leur nom ou une partie de leur nom :
  1. Cliquez sur le bouton 🔍.
  2. Dans le champ affiché, entrez le texte à rechercher.
  3. Cliquez sur le bouton > situé à droite du champ de texte pour sélectionner les attributs.

  Par exemple, si un jeu d'attributs utilise le même préfixe, vous pouvez sélectionner tous les attributs en entrant le préfixe dans le champ de texte.
1. Cliquez sur le bouton > situé entre les listes pour ajouter les attributs à la liste *Attributs sélectionnés*.
- Pour ajouter tous les attributs disponibles à la liste *Attributs sélectionnés* :
  - Cliquez sur le bouton >>.

5.10.3.3.3 Info-bulles

Les info-bulles vous permettent d'afficher des informations supplémentaires sur les nœuds lorsque vous faites passer le curseur sur un nœud. Les info-bulles sont activées par défaut et tous les attributs disponibles sont affichés.

- Pour désactiver les info-bulles :
  1. Dans la liste située à gauche, sélectionnez l'option *Info-bulles*.
  2. Décochez la case *Activer les info-bulles*.
- Pour modifier les attributs affichés dans les info-bulles :
  - Dans la liste *Attributs sélectionnés*, sélectionnez les attributs que vous ne souhaitez pas afficher dans les info-bulles.
    - Pour sélectionner et retirer plusieurs attributs simultanément :
      1. Maintenez la touche `Ctrl` enfoncée.
      2. Sélectionnez les attributs à ajouter.
      3. Cliquez sur le bouton < pour retirer les attributs de la liste *Attributs sélectionnés*.
    - Pour trouver et retirer des attributs par leur nom ou une partie de leur nom :
      1. Cliquez sur le bouton 🔍.
      2. Dans le champ affiché, entrez le texte à rechercher.
      3. Cliquez sur le bouton > situé à droite du champ de texte pour sélectionner les attributs.

    Par exemple, si un jeu d'attributs utilise le même préfixe, vous pouvez sélectionner tous les attributs en entrant le préfixe dans le champ de texte.
    1. Cliquez sur le bouton **<** situé entre les listes pour retirer les attributs de la liste *Attributs sélectionnés*.
    - Pour retirer tous les attributs de la liste *Attributs sélectionnés* :
      - Cliquez sur le bouton <<.

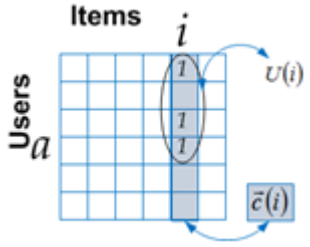
5.10.3.4 Recommandations

Ce panneau vous permet de générer des recommandations à partir d'un graphe projeté d'un graphe bipartite. Par exemple, si vous avez créé un graphe bipartite reliant des clients et des produits et avez projeté un graphe reliant les produits, vous pouvez générer des recommandations produits pour un client particulier.

- Pour paramétrer les options
  1. Dans la liste *Moteur de projection*, sélectionnez le graphe contenant les éléments que vous voulez recommander.
  2. Vous pouvez personnaliser votre jeu de données de sortie en paramétrant les options suivantes :
  - *Ne pas recommander si déjà possédé*
    Par défaut les produits déjà achetés par le client ne seront pas recommandés.
    Décochez cette option si vous voulez inclure ces produits dans les recommandations.
  - *Inclure les meilleures ventes*
    Par défaut, les meilleures ventes ne sont pas incluses dans les recommandations. La recommandation des produits les mieux vendus n'est pas liée aux clients mais provient d'une analyse spécifique.
    Cochez cette option si vous souhaitez que les meilleures ventes soient incluses dans les produits recommandés.
  - *UserID*
    Le nom de ce champ dépend du *Moteur de projection*. Il s'agit du nom de la variable contenant l'identifiant de l'élément pour lequel vous voulez créer des recommandations.
    Saisissez dans ce champ la valeur correspondant à cet élément.
  - *Garder les N plus importants*
    Par défaut, seules les cinq meilleures recommandations sont fournies dans le jeu de données de résultats.
    Pour modifier le nombre de recommandations fournies dans les résultats, changez le nombre indiqué dans le champ *Garder les N plus importantes*.
    Pour obtenir la liste complète des recommandations générées, décochez cette option. Attention, il peut y avoir un grand nombre de recommandations pour un client, il faut donc être prudent lorsque vous désactivez cette option.
  - *Baser les recommandations sur*
    Cette option vous permet de sélectionner la métrique à utiliser pour trier les recommandations. Par défaut, les recommandations sont triées par ordre décroissant de confiance. (Support, Confiance, KI, Confiance combinée, Cosinus)
  - *Utiliser la somme*
    Lorsque cette option est sélectionnée, le critère de tri est la somme de la métrique de toutes les règles qui recommande ce produit particulier, c'est-à-dire qui ont le même conséquent.
  - *Utiliser le poids*
    Cette option n'est utilisable qu'avec l'option *Utiliser la somme*. Elle permet d'utiliser la colonne de poids définie dans les paramètres du modèle pour pondérer la somme.
  - *Facteur de combinaison*
    Cette option permet de paramétrer la confiance combinée - voir *Pour générer des recommandations* ci-dessous.
- Pour générer des recommandations

Cliquez sur le bouton *Accéder aux recommandations*.

Une table répertoriant les éléments recommandés s'affiche. Elle fournit les informations suivantes :

Intitulé	Description
Recommandation	Élément recommandé (nœud)
Support	Support de la règle utilisée pour recommander cet élément
Confiance	Confiance de la règle utilisée pour recommander cet élément
KI	Capacité prédictive (KI) de la règle utilisée pour recommander cet élément
Confiance combinée	$Sim(i,j) = P(i/j)^{\alpha} P(j/i)^{1-\alpha}$ où α est le Facteur de combinaison. Lorsque que le facteur de combinaison est égal à 1, la confiance combinée est égale à la confiance. Lorsque que le facteur de combinaison est égal à 0,5, la confiance combinée est égale au cosinus.
Cosinus	 où a : l'utilisateur actif i et j : deux éléments $U(i)$: le nombre d'utilisateurs achetant i $\vec{c}(i)$: la colonne de la matrice R pour l'élément i $Sim(i,j) = cos[\vec{c}(i), \vec{c}(j)] = \frac{\vec{c}(i) \bullet \vec{c}(j)}{\|\vec{c}(i)\|_2 \times \|\vec{c}(j)\|_2} = \frac{\sum_{u=1}^{L} w_{ui} \times w_{uj}}{\sqrt{\sum_{u=1}^{L} (w_{ui})^2} \times \sqrt{\sum_{u=1}^{L} (w_{uj})^2}}$
Rôle	Il existe quatre types de rôles dans les communautés : - Nœud passerelle : nœud reliant des communautés entre elles en agissant comme un pont. Ce type de nœud comprend beaucoup de liens avec l'extérieur de sa communauté, mais peu à l'intérieur de celle-ci. - Nœud local : nœud comprenant un grand nombre de liens dans une communauté, mais très peu de liens avec d'autres communautés. - Nœud social : nœud assurant la cohésion des communautés par leur grand nombre de liens dans leur propre communauté mais également vers d'autres communautés. - Nœud passif : nœud comprenant peu de liens à l'intérieur comme à l'extérieur de sa communauté. Ce type de nœud se trouve plutôt en périphérie de la communauté.

Intitulé	Description
Id de communauté	Sous-graphe dont la densité des connexions internes est supérieure à celle de ses connexions vers le reste des nœuds du réseau.

5.10.3.5 Recommandations en masse

Lors de la création de recommandations avec le module Social, ce panneau vous permet de générer une liste des produits recommandés pour chaque client de votre jeu de données.

- Paramétrage des options de génération [page 257]
- Sélection du jeu de données d'application [page 259]
- Enregistrement des résultats générés par le modèle [page 260]
- Mode Application in-database [page 259]

5.10.3.5.1 Paramétrage des options de génération

- Pour paramétrer les options de génération :
  1. Dans la liste *Recommandeur*, sélectionnez le recommandeur que vous souhaitez utiliser pour générer les recommandations.
  2. Vous pouvez personnaliser votre jeu de données de sortie en paramétrant les options suivantes :
     - *Ne pas recommander si déjà possédé*
       Par défaut les produits déjà achetés par le client ne seront pas recommandés.

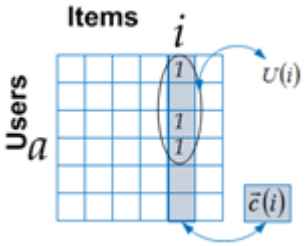
     Décochez cette option si vous voulez inclure ces produits dans les recommandations.
     - *Inclure les meilleures ventes*
       Par défaut, les meilleures ventes ne sont pas incluses dans les recommandations. La recommandation des produits les mieux vendus n'est pas liée aux clients mais provient d'une analyse spécifique.

     Cochez cette option si vous souhaitez que les meilleures ventes soient incluses dans les produits recommandés.
     - *Garder les N plus importants*
       Par défaut, seules les cinq meilleures recommandations sont fournies dans le jeu de données de résultats.
       - Pour modifier le nombre de recommandations fournies dans les résultats, changez le nombre indiqué dans le champ *Garder les N plus importantes*.
       - Pour obtenir la liste complète des recommandations générées, décochez cette option. Attention, il peut y avoir un grand nombre de recommandations pour un client, il faut donc être prudent lorsque vous désactivez cette option.
     - *Baser les recommandations sur*
       Cette option vous permet de sélectionner la métrique à utiliser pour trier les recommandations. Par défaut, les recommandations sont triées par ordre décroissant de confiance.
     - *Utiliser la somme*

Lorsque cette option est sélectionnée, le critère de tri est la somme de la métrique de toutes les règles qui recommande ce produit particulier, c'est-à-dire qui ont le même conséquent.

- *Utiliser le poids*
  Cette option n'est utilisable qu'avec l'option *Utiliser la somme*. Elle permet d'utiliser la colonne de poids définie dans les paramètres du modèle pour pondérer la somme.
- *Facteur de combinaison*
  Cette option permet de paramétrer la *Confiance combinée* - voir Métriques et formules ci-dessous.

Métriques et formules

Métriques	Formules et notes
Confiance combinée	$Sim(i,j) = P(i/j)^{\alpha} P(j/i)^{1-\alpha}$ où α est le facteur de combinaison. • Lorsque que le facteur de combinaison est égal à 1, la confiance combinée est égale à la confiance. • Lorsque que le facteur de combinaison est égal à 0,5, la confiance combinée est égale au cosinus.
Cosinus	 où • a : l'utilisateur actif • i et j : deux éléments • $U(i)$: le nombre d'utilisateurs achetant i • $\vec{c}(i)$: la colonne de la matrice R pour l'élément i $Sim(i,j) = cos[\vec{c}(i),\vec{c}(j)] = \frac{\vec{c}(i) \bullet \vec{c}(j)}{\|\vec{c}(i)\|_2 \times \|\vec{c}(j)\|_2} = \frac{\sum_{u=1}^{L} w_{ui} \times w_{uj}}{\sqrt{\sum_{u=1}^{L} (w_{ui})^2} \times \sqrt{\sum_{u=1}^{L} (w_{uj})^2}}$
KI	$Sim(i,j) = (P(i \,\&\, j) - P(i) * P(j)) / (1 - P(i)) * P(i)$

5.10.3.5.2 Sélection du jeu de données d'application

- Pour sélectionner le jeu de données d'application
  1. Dans la section *Jeu de données d'application*, cliquez sur le bouton *Parcourir* sur la ligne *Dossier* pour sélectionner le sous-répertoire du fichier source.
  2. Cliquez sur le bouton *Parcourir* sur la ligne *Données* pour sélectionner le fichier texte ou la source ODBC. Le format des données d'entrée doit être identique à celui utilisé pour créer le modèle.

> **i** Remarque
>
> - Un message d'avertissement peut s'afficher indiquant que le jeu de données contient plus de champs que déclaré dans le modèle. Il s'agit d'un simple message d'information et ne requiert aucune action de votre part. Cliquez sur *Fermer* pour poursuivre.
> - Si vous avez activé le *Mode communauté* lors du paramétrage du modèle, le jeu de données d'application doit inclure la variable kxComIndex contenant les identifiants de communautés. Il doit s'agir d'un entier nominal.
>   Cette variable vous permet d'appliquer le modèle sur les graphes de communautés (détection des communautés voisines, liste des nœuds d'une communauté, statistiques agrégées, ...) Vous pouvez utiliser cette colonne pour fournir une liste d'identifiants de communautés pour lesquelles vous souhaitez effectuer des calculs. Si vous n'appliquez pas le modèle sur un graphe de communauté, cette variable ne sera pas utilisée, vous pouvez donc laissez cette colonne vide ou, dans le cas d'une base de données, utiliser une valeur factice.

5.10.3.5.3 Mode Application in-database

Ce mode optimisé permet d'appliquer le modèle directement dans la base de données. Cela évite d'avoir à extraire les données de la base de données et accélère le processus d'écriture des données de sortie du modèle. Ce mode peut être utilisé si toutes les conditions suivantes sont remplies :

- Le jeu de données d'application d'entrée (table, vue, requête, manipulation de données) et le jeu de données de résultat sont des tables de la même base de données.
- Le jeu de données d'entrée utilisé pour créer le modèle contient au moins une variable déclarée comme clé.
- Le mode *Application in-database* est activé.
- Vous avez les droits de lecture et d'écriture dans la base de données.
- Vous possédez une licence de générateur de codes valide pour la base de données.
- Aucune erreur ne s'est produite.
- Le modèle a été enregistré dans la base de données qui contient les jeux de données d'application.

> **i** Remarque
>
> Si le modèle n'a pas été enregistré, une boîte de dialogue s'affiche vous proposant de l'enregistrer.

5.10.3.5.4 Enregistrement des résultats générés par le modèle

- Pour enregistrer les résultats générés par le modèle :
  - Dans la section *Résultats générés par le modèle*, sélectionnez le type du jeu de données de sortie dans la liste *Type de données*.
  - Cliquez sur le bouton *Parcourir* sur la ligne *Dossier* pour spécifier le répertoire qui contiendra les résultats.
  - Saisissez un nom pour le fichier de sortie dans le champ *Données*, ou utilisez le bouton *Parcourir* pour sélectionner un fichier.
  - Cliquez sur *Appliquer*.

> **i** Remarque
>
> Un message d'avertissement peut s'afficher indiquant que le jeu de données contient plus de champs que déclaré dans le modèle. Il s'agit d'un simple message d'information et ne requiert aucune action de votre part. Cliquez sur *Fermer* pour poursuivre.

5.10.4 Géolocalisation : Co-emplacement et chemin

5.10.4.1 Paramètres pour transactions

Ce panneau vous permet de définir les paramètres dont l'application a besoin pour extraire les co-localisations ou les chemins existant dans vos données.

Paramètres obligatoires

Les paramètres suivants doivent être définis :

Paramètre	Description
Élément géolocalisé	Sélectionnez la variable contenant l'identifiant de l'élément géolocalisé pour lequel vous souhaitez trouver les co-localisations ou les chemins. Remarque : Seules les variables nominales sont visibles dans la liste. La première est sélectionnée par défaut.

Paramètre	Description
Position	Sélectionnez la variable contenant la position de l'élément géolocalisé, à savoir la variable composite de latitude et de longitude créée lors de la description des données. Remarque : Seules les variables composites de type position sont visibles dans la liste. La première est sélectionnée par défaut.
Dimension du pavage en mètres	Entrez la dimension du pavage à utiliser. Les tuiles sont générées par l'application pour définir l'emplacement d'un élément. Deux événements ayant lieu dans la même tuile sont considérés comme étant sur le même emplacement.
Colonne de date	Sélectionnez la variable contenant la date à laquelle l'élément a été géolocalisé. Remarque : Seules les variables de type date et datetime sont visibles dans la liste. La première est sélectionnée par défaut.

Période

Cette section vous permet de filtrer les événements sur une période de temps déterminée. Vous pouvez définir uniquement une date de début, uniquement une date de fin ou les deux.

Paramètres	Description
Date de début	Prend en compte uniquement les événements qui se produisent après une date spécifique. La date de début est incluse dans la période.
Date de fin	Prend en compte uniquement les événements qui se produisent avant une date spécifique. La date de fin est exclue de la période.

Divers

Paramètre	Description
Nombre maximal de connexions	Nombre maximal de connexions pour un nœud (emplacement). Les nœuds contenant plus de connexions que le maximum autorisé ne sont pas pris en compte lors de l'entraînement du modèle.

5.10.4.1.1 Sélection du jeu de données

- Pour sélectionner un jeu de données
  1. Dans la liste *Type de données*, sélectionnez le type de données auquel vous souhaitez accéder. L'accès à certains types de données nécessite une licence spécifique.
  2. Utilisez le bouton *Parcourir* correspondant au champ *Dossier* pour sélectionner le dossier ou la base de données contenant les données désirées. En cas de *base de données protégée*, vous devrez saisir le *nom d'utilisateur* et le *mot de passe* lors de la sélection de la base de données.
     Lorsque le format des données le permet, vous disposez des options suivantes dans la fenêtre *Sélection de données* :

     : aller au répertoire parent

     : créer un dossier

     : supprimer le dossier sélectionné (cela n'est possible que pour des dossiers vides)
  3. Utilisez le bouton *Parcourir* correspondant au champ *Événements* pour sélectionner le jeu de données à utiliser. Les jeux de données créés d'analyse créés avec SAP Predictive Analytics sont affichées en tête de liste et indiqués par l'icône .
  4. Le bouton Loupe permet de visualiser les données et de calculer les statistiques du jeu de données sélectionné.
  5. Le bouton *Paramètres avancés* vous permet de définir certains paramètres avancés tels que l'échantillonnage ou les paramètres de fichiers.

5.10.4.1.1.1 Sélection d'une vue SAP HANA

Vous pouvez choisir une vue analytique ou de calcul SAP HANA en tant que source de données pour votre modèle.

1. Dans l'écran *Données de transactions*, sélectionnez *Base de données* dans la liste *Type de données*.
2. Cliquez sur le bouton *Parcourir* à gauche du champ *Dossier*, puis sélectionnez la base de données SAP HANA que vous souhaitez utiliser dans la liste *Sélectionnez une source pour les données*.
3. S'il s'agit de la première connexion à cette base de données, saisissez les informations de connexion dans les champs *Utilisateur* et *Mot de passe* , puis cliquez sur *Se connecter*.

   Le contenu de la base de données s'affiche dans la partie gauche de la fenêtre.
4. Ouvrez le nœud *Contenu*. Utilisez l'arborescence pour parcourir les packages et sélectionnez la vue SAP HANA que vous souhaitez utiliser en tant que source de données pour votre modèle.
5. Cliquez sur *OK*.

   Si la vue SAP HANA que vous avez sélectionnée requiert la définition de certaines valeurs, la fenêtre *Valeurs utilisateur* s'ouvre lorsque vous cliquez sur *Analyser*. Les variables pour lesquelles vous devez saisir une valeur sont répertoriées à gauche du panneau. Le tableau suivant présente en détails la méthode d'identification des différentes variables dans la liste. Dans cette rubrique, le terme variable est utilisé à la fois pour les variables et les paramètres d'entrée.

Type de variable	Noté comme	Signalé par
Variable obligatoire	obligatoire	Une croix rouge
Variable facultative	facultative	
Variable avec une valeur par défaut	valeur par défaut	Une coche verte
Variable avec une valeur définie par l'utilisateur		Une coche verte

Cochez la case *Afficher uniquement obligatoires* pour afficher uniquement les variables obligatoires dans la liste.

6. Dans la liste, sélectionnez les variables pour lesquelles vous souhaitez définir les valeurs.

   Pour cela, le champ approprié est affiché à droite du panneau. Selon le type de valeur requis, il est possible que plusieurs types de champs s'affichent. Par exemple, si une valeur unique est requise, une liste des opérateurs et un champ de saisie de valeur s'affichent. Si une plage de valeurs est requise, vous verrez apparaître deux champs pour les limites de plage et une liste des opérateurs.
7. Certaines variables contiennent une liste de valeurs possibles prédéfinies. Dans ce cas, la liste des valeurs pour la variable actuelle est affichée dans la partie droite du panneau. Sélectionnez la valeur que vous souhaitez définir pour cette variable ; elle sera affichée dans le champ de valeur.
8. Utilisez le bouton *Restaurer valeur par défaut* pour revenir à la valeur par défaut de la variable en cours.
9. Une fois toutes les valeurs requises saisies, cliquez sur *OK*.

   Après avoir accédé à la vue SAP HANA, les valeurs définies seront mémorisées et utilisées par défaut ultérieurement lors du processus de modélisation.

5.10.4.1.2 Description des données

Il existe deux façons de décrire les données :

- en analysant le jeu de données ;
- en se servant d'une description du jeu de données complet préalablement enregistrée sous forme de fichier ou de table de base de données.

5.10.4.1.3 Analyse du jeu de données

L'analyse du jeu de données d'entraînement vous permettra de commencer avec une description obtenue soit par l'analyse des cents premières lignes de données, soit par l'utilisation d'une description issue de la bibliothèque de variables si une variable de même nom s'y trouve. Vous pouvez ensuite modifier les descriptions proposées (en vérifiant surtout le genre des variables).

> **i** Remarque
>
> Pour les fichiers texte, il est bon d'avoir une première ligne d'en-tête contenant les noms des variables.

Lorsque vous créez la description par *analyse* automatique, vérifiez chaque variable afin de valider la description comme indiqué dans la définition Description d'une variable.

- Pour analyser le jeu de données
  - Cliquez sur le bouton *Analyser* situé dans l'onglet *Accueil* du ruban.
    Les variables dont la description provient de la bibliothèque de variables sont signalées par l'icône dans la colonne *Index*.

5.10.4.1.4 Chargement d'une description du jeu de données

Pour décrire le jeu de données, vous pouvez utiliser un fichier ou une table de base de données défini(e) précédemment et contenant les descriptions de toutes les variables.

- Pour ouvrir une description de jeu de données
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Ouvrir une description*.
  2. Dans la fenêtre *Ouvrir une description pour...*, sélectionnez le format de la description dans la liste déroulante *Type de données*.
  3. Utilisez le bouton *Parcourir* à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où doit être stockée la description.
  4. Utilisez le bouton *Parcourir* situé à droite du champ *Description* pour sélectionner le fichier ou la table contenant la description.
  5. Cliquez sur *OK*. La fenêtre se ferme et la description s'affiche dans le panneau principal. Vous pouvez alors la modifier si nécessaire.

5.10.4.1.5 Ajout d'une variable composite pour la position

Lorsque vous décrivez les données, vous devez définir une variable composite de type *position*.

Assurez-vous que les variables de latitude et de longitude de votre jeu de données ont un type de stockage *angle*.

- Pour ajouter des variables composites :
  1. Accédez à l'onglet *Édition* et cliquez sur *Variables Composites*.
  2. Cliquez sur *Ajouter* pour entrer le nom de la variable composite que vous souhaitez ajouter.
  3. Dans la liste *Type*, sélectionnez *Position.*
  4. Sélectionnez les variables de *Latitude* et *Longitude* dans la liste.
  5. Cliquez *OK* pour valider.
- Pour modifier une variable composite
  1. Cliquez sur *Variables Composites*.
  2. Sélectionnez celle que vous souhaitez modifier.
  3. Cliquez sur *Modifier*.
- Pour supprimer une variable composite
  1. Dans *Variables Composites*, sélectionnez la variable composite que vous souhaitez supprimer.
  2. Cliquez sur *Supprimer*.

5.10.4.1.6 Enregistrement d'une description

Vous pouvez enregistrer la description des variables soit dans la bibliothèque de variables, soit dans un fichier ou une table de base de données.

5.10.4.1.7 Dans la bibliothèque de variables

Vous pouvez choisir d'enregistrer les descriptions de toutes les variables, ou sélectionner uniquement celles qui vous intéressent. Vous pouvez aussi supprimer des variables de la bibliothèque de variables.

Si vous avez défini un référentiel de métadonnées unique (option accessible via le bouton *Aide*, puis *Modifier les options...*, onglet *Référentiel de métadonnées*), les descriptions sont enregistrées dans ce référentiel sinon elles sont enregistrées au même emplacement que le jeu de données que vous décrivez.

- Pour enregistrer toutes les descriptions :
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Enregistrer dans la bibliothèque de variables*.
  2. Sélectionnez l'option *Toutes les variables*.
- Pour enregistrer uniquement certaines descriptions :
  1. Sélectionnez la variable dont vous voulez enregistrer la description. Pour sélectionner plusieurs variables, utilisez la touche `Ctrl`.
  2. Cliquez sur le bouton *Enregistrer dans la bibliothèque de variables*.
  3. Sélectionnez l'option *Variables sélectionnées*.
- Pour supprimer de la bibliothèque de variables :
  - Sélectionnez la variable à supprimer de la bibliothèque de variables et cliquez sur *Supprimer* de la bibliothèque de variables.

5.10.4.1.8 Dans un fichier ou une table de base de données

Vous pouvez enregistrer l'ensemble des descriptions du jeu de données dans un fichier ou dans une table de base de données.

- Pour enregistrer la description du jeu de données dans un fichier ou une table :
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Enregistrer la description*. La fenêtre *Enregistrer une description de...* s'affiche.
  2. Dans la liste *Type de données*, sélectionnez le format dans lequel vous voulez enregistrer la description.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour choisir le dossier ou la base de données où vous souhaitez enregistrer la description.
  4. Dans le champ *Description*, saisissez le nom du fichier ou de la table où vous souhaitez enregistrer la description.
  5. Cliquez sur *OK*.

5.10.4.2 Co-localisation

Ce panneau vous permet de définir les paramètres des co-localisations.

Pour créer une co-localisation basée sur une co-localisation déjà définie, cliquez sur *Cloner*.

Paramètres des co-localisations

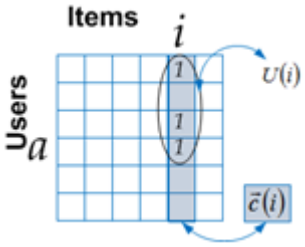
Pour définir les co-localisations, les valeurs suivantes doivent être paramétrées :

Paramètre	Description
Nom de la co-localisation	Saisissez le nom de la co-localisation. **i Remarque** Le nom ne doit pas contenir les caractères suivants : barre oblique (/), supérieur à (>), inférieur à (<).
Intervalle de co-localisations	Saisissez l'intervalle maximal entre deux événements considérés comme simultanés. Par exemple, si l'*Intervalle de co-localisation* est défini sur 15 minutes, cela signifie que deux événements sont co-localisés, s'ils se produisent dans un intervalle de 15 minutes.
Nombre maximal de co-localisations	Sélectionnez l'option *Nombre maximal de co-localisations* pour limiter le nombre de co-localisations conservées dans le modèle par l'application. Cela limite l'utilisation de la mémoire et simplifie l'analyse. Si cette limite est dépassée, l'application conserve les co-localisations ayant le plus de connexions.

Paramètres avancés

La table ci-dessous liste tous les paramètres avancés disponibles :

Paramètre	Description
Support minimal	Le support est le nombre d'éléments que deux entités ont en commun. Ainsi, lorsque vous projetez un graphe bipartite reliant des produits et des clients, le support du lien joignant le produit A et le produit B est le nombre de clients ayant acheté les deux produits. Un lien dont le support est inférieur au *Support minimal* ne sera pas créé.
Confiance minimale	La confiance d'un lien est le pourcentage d'enregistrements qui vérifient le conséquent de la règle parmi ceux qui vérifient son antécédent. Dans le cas d'une règle de la forme X=>Y, la confiance de la règle est le nombre d'enregistrements contenant à la fois le jeu d'éléments X et l'attribut Y comparé au nombre d'enregistrements contenant uniquement le jeu d'éléments X. Si vous n'avez pas coché la case *Dirigé*, et même si la confiance d'un lien n'est supérieure à la *Confiance minimale* que pour une des directions, le lien sera conservé.
KI minimal	Seuls les liens pour lesquels le KI est supérieur au KI minimal seront conservés. Si vous n'avez pas coché la case *Dirigé*, même si le KI d'un lien n'est supérieur au *KI minimal* que pour une des directions, le lien sera conservé.
Dirigé	Filtre permettant de différencier les liens en fonction de leur direction et de fournir des règles dirigées.
Limite de transition suivante	Limite le nombre de transitions prises en compte par l'application après qu'une paire soit trouvée.
Garder les N plus importants	Le nombre d'appariement de nœuds à garder parmi les mieux classés.

Paramètre	Description
Poids	Indique quelle valeur utiliser comme poids pour les liens. Il est recommandé d'utiliser l'indice de Jaccard ou le Ratio d'indépendance, ce dernier en particulier lorsque vous travaillez avec des communautés. • Support Utilise le support tel que défini ci-dessus. • Indice de Jaccard Permet d'évaluer la similarité entre des ensembles. Il est défini comme la taille de l'intersection divisée par la taille de l'union des ensembles : $J(A,B) = \frac{\lvert A \cap B \rvert}{\lvert A \cup B \rvert}$ • Ratio d'indépendance Deux événements A et B sont indépendants si la probabilité que les deux événements se produisent est égale à la probabilité de l'événement A fois la probabilité de l'événement B : $\Pr(A \cap B) = \Pr(A).\Pr(B)$ Le poids est alors calculé suivant la formule ci-dessous : $w = \frac{\Pr(A \cap B)}{\Pr(A).\Pr(B)}$ Un poids de 1 indique deux événements complètement indépendants. Plus le poids est élevé, plus les événements sont corrélés. • Cosinus  où a : l'utilisateur actif i et j : deux éléments : le nombre d'utilisateurs achetant i : la colonne de la matrice R pour l'élément i

Paramètre	Description
	$Sim(i,j) = cos[\vec{c}(i), \vec{c}(j)] = \frac{\vec{c}(i) \bullet \vec{c}(j)}{\|\vec{c}(i)\|_2 \times \|\vec{c}(j)\|_2} = \frac{\sum_{u=1}^{L} w_{ui} \times w_{uj}}{\sqrt{\sum_{u=1}^{L} (w_{ui})^2} \times \sqrt{\sum_{u=1}^{L} (w_{uj})^2}}$
Privilégier la consommation de la mémoire	Cette option ralentit le processus mais consomme moins de mémoire.
Privilégier la vitesse de traitement	Cette option accélère le processus mais consomme plus de mémoire.

5.10.4.3 Extraction de chemin

Ce panneau vous permet de définir les paramètres de chemin.

Pour créer une extraction de chemin basée sur une extraction déjà définie, cliquez sur *Cloner*.

Paramètres de chemin

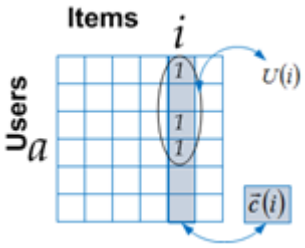
Pour définir les chemins à extraire, les valeurs suivantes doivent être paramétrées :

Paramètre	Description
Nom de séquence	Saisissez le nom de l'extraction de chemin. **i Remarque** Le nom ne doit pas contenir les caractères suivants : barre oblique (/), supérieur à (>), inférieur à (<).
Durée maximale entre les transitions	Saisissez l'intervalle maximal entre deux événements considérés comme appartenant au même chemin. Par exemple, si la *Durée minimale entre les transitions* est définie sur 15 minutes, un élément doit atteindre et quitter le même emplacement dans un intervalle de 15 minutes, sinon, la chaîne des événements (chemin) est rompue.
Nombre de chemins maximal	Sélectionnez l'option *Nombre de chemins maximal* pour limiter le nombre de chemins conservés dans le modèle par l'application. Cela limite l'utilisation de la mémoire et simplifie l'analyse.
Longueur de chemin maximale	Vous pouvez modifier le nombre de segments dans le chemin (un segment correspond à la partie du chemin entre deux emplacements). L'allongement de la longueur de chemin entraine un allongement de la durée d'analyse.

Paramètres avancés

La table ci-dessous liste tous les paramètres avancés disponibles :

Paramètre	Description
Support minimal	Le support est le nombre d'éléments que deux entités ont en commun. Ainsi, lorsque vous projetez un graphe bipartite reliant des produits et des clients, le support du lien joignant le produit A et le produit B est le nombre de clients ayant acheté les deux produits. Un lien dont le support est inférieur au *Support minimal* ne sera pas créé.
Confiance minimale	La confiance d'un lien est le pourcentage d'enregistrements qui vérifient le conséquent de la règle parmi ceux qui vérifient son antécédent. Dans le cas d'une règle de la forme X=>Y, la confiance de la règle est le nombre d'enregistrements contenant à la fois le jeu d'éléments X et l'attribut Y comparé au nombre d'enregistrements contenant uniquement le jeu d'éléments X. Si vous n'avez pas coché la case *Dirigé*, et même si la confiance d'un lien n'est supérieure à la Confiance minimale que pour une des directions, le lien sera conservé.
KI minimal	Seuls les liens pour lesquels le KI est supérieur au KI minimal seront conservés. Si vous n'avez pas coché la case *Dirigé*, même si le KI d'un lien n'est supérieur au *KI minimal* que pour une des directions, le lien sera conservé.
Dirigé	Filtre permettant de différencier les liens en fonction de leur direction et de fournir des règles dirigées.
Limite de transition suivante	Limite le nombre de transitions prises en compte par l'application après qu'une paire soit trouvée.
Garder les N plus importants	Le nombre d'appariement de nœuds à garder parmi les mieux classés.

Paramètre	Description
Poids	Indique quelle valeur utiliser comme poids pour les liens. Il est recommandé d'utiliser l'*indice de Jaccard* ou le *Ratio d'indépendance*, ce dernier en particulier lorsque vous travaillez avec des communautés. *Support* Utilise le support tel que défini ci-dessus. *Indice de Jaccard* Permet d'évaluer la similarité entre des ensembles. Il est défini comme la taille de l'intersection divisée par la taille de l'union des ensembles : $$J(A,B) = \frac{\lvert A \cap B \rvert}{\lvert A \cup B \rvert}$$ *Ratio d'indépendance* Deux événements A et B sont indépendants si la probabilité que les deux événements se produisent est égale à la probabilité de l'événement A fois la probabilité de l'événement B : $$\Pr(A \cap B) = \Pr(A).\Pr(B)$$ Le poids est alors calculé suivant la formule ci-dessous : $$w = \frac{\Pr(A \cap B)}{\Pr(A).\Pr(B)}$$ Un poids de 1 indique deux événements complètement indépendants. Plus le poids est élevé, plus les événements sont corrélés. *Cosinus*  où a : l'utilisateur actif i et j : deux éléments : le nombre d'utilisateurs achetant i : la colonne de la matrice R pour l'élément i

Paramètre	Description
	$Sim(i,j) = cos[\vec{c}(i), \vec{c}(j)] = \frac{\vec{c}(i) \bullet \vec{c}(j)}{\|\vec{c}(i)\|_2 \times \|\vec{c}(j)\|_2} = \frac{\sum_{u=1}^{L} w_{ui} \times w_{uj}}{\sqrt{\sum_{u=1}^{L} (w_{ui})^2} \times \sqrt{\sum_{u=1}^{L} (w_{uj})^2}}$
Privilégier la consommation de la mémoire	Cette option ralentit le processus mais consomme moins de mémoire.
Privilégier la vitesse de traitement	Cette option accélère le processus mais consomme plus de mémoire.

5.10.4.4 Post-traitement

Ce panneau vous permet de sélectionner le type de post-traitement que vous souhaitez appliquer sur les graphes que vous souhaitez créer.

- Post-traitement générique [page 244]
- Appariement des nœuds [page 246]

5.10.4.4.1 Post-traitement générique

Détection de communautés

Détection de communautés

Cette option permet d'identifier les communautés existant dans le jeu de données. Une communauté peut être considérée comme un sous-graphe dont la densité des connexions internes est supérieure à celle de ses connexions vers le reste des nœuds du réseau. La détection des communautés utilise un algorithme itératif dont les critères d'arrêt sont le **nombre maximum d'itérations** et le **gain de modularité d'une itération sur l'autre**, représenté par Epsilon.

Pour en savoir plus sur le calcul de la détection de communautés, voir les informations sur l'algorithme de Louvain .

Vous pouvez également **baser les communautés d'un graphe sur celles d'un autre graphe**. Ainsi, les communautés du mois de février peuvent être initialisées sur celle de janvier, vous permettant de visualiser leur évolution d'un mois sur l'autre.

- Pour activer la détection des communautés
  - Cochez la case **Détection des communautés** correspondant au graphe pour lequel vous voulez activer la détection.
    Remarque : La détection des communautés n'est pas disponible pour les graphes transactionnels.

- Pour configurer la détection des communautés :
  1. Cliquez sur le bouton **...** situé à droite de la case à cocher **Détection des communautés**. La fenêtre **Configuration de la détection des communautés** s'affiche.
  2. Cochez la case **Activer la détection des communautés**.
  3. Définissez les options que vous souhaitez appliquer. Le tableau suivant résume les options disponibles :

Option	Description
Nombre maximum d'itérations	Nombre maximum d'itérations avant que l'algorithme ne s'arrête si les critères Epsilon ne sont pas atteints.
Epsilon	Valeur en-dessous de laquelle le gain de modularité n'est plus suffisamment intéressant pour justifier une autre itération de l'algorithme.
Se baser sur les communautés de	graphe à partir duquel vous pouvez utiliser les communautés comme base pour le graphe en cours. La liste Se baser sur les communautés de comprend uniquement les graphes pour lesquels la détection des communautés a été activée. La base de communautés n'est pas disponible pour les graphes transactionnels.
Ignorer les auto-liens	Réduit le nombre de communautés et apporte une meilleure perspective de la structure du graphe.
Désactiver les supercommunautés	Désactive la création de supercommunautés, ce qui accélère la création lors de l'utilisation de grands graphes.

  4. Cliquez sur **OK** pour valider les paramètres.

Optimisation de la détection des méga-hubs

L'option *Optimiser la détection des méga-hubs* supprimera tous les nœuds dont le degré de connectivité est supérieur au seuil défini. Il existe deux façons de définir ce seuil :

- Utiliser le seuil automatique, basé sur la distribution des nœuds en fonction de leur degré de connectivité. Le seuil automatique est calculé à l'aide de la formule suivante : moyenne + facteur de l'écart-type * écart-type

> **i** Remarque
>
> Par défaut, le facteur de l'écart-type est fixé à 4.

- Utiliser une valeur fixe.
- Pour activer l'optimisation de la détection des méga-hubs
  - Cochez la case *Optimiser la détection des méga-hubs*.
- Pour paramétrer la détection des méga-hubs

> **i Remarque**
>
> Lorsque vous paramétrez cette option pour un graphe transactionnel, vous devez configurer le seuil pour les deux populations.

1. Cliquez sur le bouton ... situé à droite de la case à cocher *Optimisez la détection des méga-hubs*. La fenêtre *Configuration de la détection des méga-hubs* s'affiche.
2. Cochez la case *Optimiser la détection des méga-hubs*.
3. Sélectionnez le type de seuil que vous souhaitez utiliser :
   - *Seuil automatique*
   - *Seuil manuel*
4. Selon le type de seuil sélectionné :
   - pour le *seuil automatique*, entrez une valeur dans le champ *Facteur de l'écart-type* pour paramétrer la formule utilisée lors du calcul ;
   - utilisez le champ *Seuil d'exclusion* pour définir le *seuil manuel*.
5. Cliquez sur *OK* pour valider les paramètres.

5.10.4.4.2 Appariement des nœuds

L'appariement des nœuds peut être utilisé pour identifier un élément (client, produit, etc.) présent dans deux graphes (avec deux nœuds distincts) grâce à leurs voisins communs. Pour une compagnie téléphonique, cette technique peut notamment être utilisée pour détecter le churn rotationnel (identifier les clients qui changent fréquemment d'abonnement téléphonique afin de bénéficier des offres de bienvenue).

- Pour apparier les graphes :
  1. Sélectionnez l'onglet *Appariement*.
  2. Cliquez sur le bouton situé en haut à droite de l'onglet pour ajouter une nouvelle définition d'appariement. La fenêtre *Définition de l'appariement* s'ouvre.
  3. Sélectionnez le *Premier graphe*.
  4. Sélectionnez le *Second graphe*.
  5. Cochez l'option située à côté du champ *Nom du graphe en sortie* et saisissez un nom si vous souhaitez que les graphes appariés s'affichent comme un nouveau graphe.
  6. Dans la section *Conditions d'appariement*, saisissez un *nombre minimal de voisins communs*.
  7. Cochez l'option située à côté du champ *Proportion minimale de voisins communs* et indiquez une valeur si vous souhaitez définir cette condition.
  8. Cliquez sur *OK* pour valider la définition d'appariement. La fenêtre se ferme et le nouvel appariement s'affiche dans la liste.
  9. Si vous souhaitez définir un autre appariement, refaites les étapes 2 à 8.

Conditions d'appariement

Condition	Description

Nombre minimal de voisins communs	Deux nœuds sont appariés uniquement s'ils ont un nombre de voisins communs égal ou supérieur au nombre indiqué.
Garder les N plus importants	Seules les N paires les plus importantes sont conservées. L'importance des paires est déterminée par le *Type d'appariement*. Lorsque la valeur de ce champ est égale à 0, aucun filtre n'est appliqué. Les graphes conservent les N liens sortants ayant les poids les plus importants, donc lorsque l'utilisateur veut générer les graphes contenant la proportion et le nombre de voisins communs, il se peut que des liens vers des paires différentes soient conservés.
Proportion minimale de voisins communs	Seules les paires pour lesquelles l'indice de Jaccard est supérieur ou égal au nombre indiqué sont conservées.
Type d'appariement	• Ratio • Jaccard • Ratio d'indépendance (uniquement dans le cas d'appariements dans un même graphe) • Confiance • Coefficient de segmentation Les formules correspondant à chacun de ces types sont indiquées dans la section ci-dessous.
Proportion pondérée	La métrique utilisée pour calculer le *Type d'appariement* est pondérée avec la variable de poids déclarée lors de la définition du graphe.
Inclure le graphe du "Nombre de voisins communs"	Lors de la génération de graphe, un graphe supplémentaire est créé dans lequel les liens sont étiquetés avec le nombre de voisins communs.

Types d'appariement

- **Ratio**

Non-pondéré	$$ratio(A,B) = \frac{\#\,common_neighbors(A,B)}{(\#\,neighbors(A)+\#\,neighbors(B))/2}$$
Pondéré	$$ratio(A,B) = \frac{w_degree_common(A,B)}{(w_degree(A)+w_degree(B))}$$

- **Jaccard**

Non-pondéré	$$jaccard(A,B) = \frac{\#\,common_neighbors(A,B)}{\#\,neighbors(A)+\#\,neighbors(B)-\#\,common_neighbors(A,B)}$$

Pondéré	$jaccard(A,B) = \frac{w_degree_common(A,B)}{w_degree(A) + w_degree(B)}$

- **Ratio d'indépendance**

Non-pondéré	$independence_ratio(A,B) = \frac{P(A \& B)}{P(A)P(B)} = \frac{\#common_neighbors(A,B) * \#links}{\#neighbors(A) * \#neighbors(B)}$
Pondéré	$independence_ratio(A,B) = \frac{P(A \& B)}{P(A)P(B)} = \frac{w_degree_common(A,B) * total_weight}{w_degree(A) * w_degree(B)}$

- **Confiance**

Non-pondéré	$confidence(A,B) = \frac{P(A \& B)}{P(A)} = \frac{\#common_neighbors(A,B)}{\#neighbors(A)}$
Pondéré	$confidence(A,B) = \frac{P(A \& B)}{P(A)} = \frac{w_degree_common_from(A)to(A,B)}{w_degree(A)}$

- **Coefficient de segmentation**

Non-pondéré	$clustering(A,B) = \frac{2 * (\#triangles(A) + \#common_neighbors(A,B))}{\#neighbors(A) * (1 + \#neighbors(A))}$
Pondéré	$clustering(A,B) = \frac{2 * (\#triangles(A) + w_degree_common_from(A)to(A,B))}{w_degree(A) * (1 + w_degree(A))}$

5.10.4.5 Co-localisations

Le panneau *Co-localisations* vous permet de lister les éléments qui apparaissent au même emplacement au même moment comme un élément spécifique, ou de lister tous les éléments qui, à un moment donné, étaient positionnés dans une tuile spécifique au même moment.

Prérequis

Votre modèle doit contenir les graphes suivants :

- un graphe de proximité,
- et un graphe projeté à partir de ce graphe de proximité sur l'élément à localiser. Notez que lors de la création du graphe dérivé, la *Méthode de sélection* utilisée pour la *Liste des candidats* doit être *Synchrone*.

En fonction des informations que vous recherchez, vous aurez besoin de l'un ou des deux articles suivants :

- l'identifiant de l'élément pour lequel vous souhaitez lister les co-localisations Vous pouvez l'obtenir directement depuis votre jeu de données ou via le panneau *Afficher le nœud* accessible depuis le panneau *Options de l'analyse des réseaux sociaux*.
- l'identifiant de la tuile pour laquelle vous souhaitez lister les co-localisations Vous pouvez l'obtenir depuis la liste des co-localisations dans ce panneau, ou depuis le panneau *Afficher le nœud*. Si vous avez créé

un graphe dérivé sur les positions pour générer des séquences, vous pouvez aussi voir les tuiles et leurs identifiants dans Google Earth lors de l'affichage des séquences.

Lister les co-localisations

- Pour lister les co-localisations pour un élément
  1. Dans le champ *Projection*, sélectionnez le graphe dérivé contenant l'élément que vous voulez localiser.
  2. Dans la liste *Afficher les co-localisations*, sélectionnez *Pour un élément*.
  3. Dans le champ *Identifiant*, entrez l'identifiant de l'élément pour lequel vous souhaitez identifier les co-localisations.
  4. Cliquez sur *Lister les co-localisations*.
- Pour lister les co-localisations dans une tuile
  1. Dans le champ *Projection*, sélectionnez le graphe dérivé contenant l'élément que vous voulez localiser.
  2. Dans la liste *Afficher les co-localisations*, sélectionnez *Dans une tuile*.
  3. Dans le champ *Identifiant*, entrez l'identifiant de la tuile pour laquelle vous souhaitez lister les co-localisations.
  4. Cliquez sur *Lister les co-localisations*.

Analyse de la liste

La liste est affichée sous forme de tableau à 5 colonnes :

La colonne...	contient...
Élément 1	l'identifiant du premier élément. Si vous avez sélectionné l'option pour lister les co-localisations d'un élément, il s'agit de l'identifiant que vous avez saisi dans le champ Identifiant.
Date 1	la date à laquelle le premier élément a été localisé dans la tuile.
Élément 2	l'identifiant du deuxième élément, à savoir, l'élément localisé dans la même tuile au même moment.
Date 2	la date à laquelle le deuxième élément a été localisé dans la tuile.
Tuile	l'identifiant de la tuile sur laquelle les deux éléments ont été localisés. Si vous avez sélectionné l'option pour lister toutes les co-localisations d'une tuile spécifique, il s'agit de l'identifiant que vous avez saisi dans le champ Identifiant.

Cela signifie que l'élément 5238 a été localisé dans la tuile 14561571596000 le mardi 20 avril 2010 à 14:59 et l'élément 9832 a été localisé dans la même tuile le même jour à 15:11. Puisque dans les paramètres du modèle, un intervalle maximum de 15 minutes a été défini pour que deux éléments soient considérés dans le même

emplacement au même moment, nous pouvons considérer que les éléments 5238 et 9832 étaient dans la même tuile au même moment.

Groupement des résultats

Vous pouvez grouper les résultats pour afficher le nombre de co-localisations dans chaque tuile, ou le nombre de co-localisations pour chaque élément.

- Pour grouper des résultats
  - Sélectionnez l'option que vous voulez utiliser dans la liste *Regroupement* :
    - *Aucun* : affiche la liste détaillée des résultats
    - *Par élément* : affiche, pour chaque couple d'éléments, le nombre de fois où les deux éléments ont été trouvés dans la même tuile au même moment. Par exemple, si le tableau des résultats contient la ligne suivante :

Élément 1	Élément 2	Nombre
9832	5238	5

Cela signifie que les éléments 9832 et 5238 se sont trouvés 5 fois dans la même tuile au même moment.

    - *Par tuile* : affiche, pour chaque tuile, le nombre de co-localisations trouvées. Par exemple, si le tableau contient la ligne suivante :

Tuile	Nombre
14555605410570	1
14561571596000	5

Cela signifie que dans le jeu de résultats, un seul couple d'éléments a été trouvé dans la première tuile alors que 5 ont été trouvés dans la deuxième.

5.10.4.6 Visualisation des séquences

Le panneau *Visualisation des séquences* vous permet de sélectionner les séquences que vous souhaitez visualiser dans Google Earth.

Dans le tableau, la colonne *ID* correspond au numéro du chemin. La colonne *Séquences* correspond aux identifiants de la tuile de départ et de la tuile d'arrivée.

Filtrer les séquences

- Pour filtrer les séquences :
  1. Dans la liste *Projection*, sélectionnez le graphe contenant les données de géolocalisation à visualiser.
  2. Dans la liste *Longueur des séquences*, sélectionnez un opérateur et une valeur. L'opérateur et la valeur déterminent la longueur de votre chemin.

- L'opérateur peut être égal à, inférieur à, supérieur à ou dans l'intervalle.
- La valeur correspond au nombre de tuiles que comportera votre chemin. Plus votre chemin comportera de tuiles, et plus il sera long.

3. Cliquez sur *Filtrer séquences* pour valider ce que vous venez de définir.

- **Visualiser les séquences**

  Le bouton *Sélection* vous permet de faire des sélections spécifiques. Vous pouvez choisir :

  - d'inverser la sélection ;
  - de sélectionner toutes les lignes ;
  - de désélectionner toutes les lignes.

  Vous pourrez ensuite visualiser les chemins sélectionnées dans Google Earth.

  - Pour obtenir un aperçu des séquences :
    1. Sélectionnez les séquences que vous souhaitez visualiser.
    2. Cliquez sur *Aperçu de la sélection*. Une fenêtre s'ouvre dans Google Earth. Les chemins sont affichés avec leur latitude et longitude.

- **Exporter les séquences**

  - Pour exporter des séquences :
    1. Dans la liste déroulante **Type d'export**, sélectionnez le type de fichier vers lequel vous voulez exporter les données des séquences.

       Types de fichiers disponibles pour l'export

Type de fichier	Description
KML	KML (Keyhole Markup Language) est un langage basé sur le formalisme XML et destiné à la gestion de l'affichage de données géospatiales dans les logiciels de SIG.
GML	Le Geography Markup Language (GML) est un langage dérivé du XML pour encoder, manipuler et échanger des données géographiques.
SHAPE_FILE	Shape File, ou "fichier de formes", est un format de données vectorielles géospatiales très répandu pour les systèmes d'informations géographiques.

    2. Cliquez sur *Exporter*.
    3. Un menu s'affiche vous permettant de choisir d'exporter *toutes les séquences* ou seulement celles que vous avez préalablement *sélectionnées*.

- **SIG**

  Le panneau *Visualisation des séquences* reproduit un SIG (système d'information géographique). Un SIG (ou GIS en anglais) est un système qui permet de capturer, d'organiser, d'analyser et d'afficher tous types de données géographiques. Social vous permet d'ajouter des séquences au SIG de façon à les modifier et les afficher.

  - Pour ajouter des séquences :

  À noter que pour ajouter une séquence au SIG, vous devez définir une adresse dans le champ *Protocole du système d'information géographique* dans les options de l'application.

  1. Dans le tableau, sélectionnez les séquences de votre choix.
  2. Cliquez sur *Ajouter les séquences*. Le champ situé à côté de ce bouton est automatiquement renseigné avec l'adresse définie dans les options de l'application, accessibles depuis le menu *Fichier* > *Préférences...*, suivi des coordonnées du centre de gravité de la séquence.

3. Cliquez sur *Ouvrir*.

5.11 Recommandation

Cette section décrit comment créer et utiliser les modèles de recommandation avec SAP Predictive Analytics.

5.11.1 Définition des paramètres

5.11.1.1 Recommandeurs

Ce panneau vous permet de créer différents recommandeurs :

- *Dupliquer le recommandeur* crée une copie d'un recommandeur avec tous ses paramètres. Vous pouvez ensuite modifier les paramètres pour obtenir des résultats différents puis conserver le meilleur modèle pour votre problématique métier.
- *Supprimer le recommandeur* supprime le recommandeur sélectionné.

Description des paramètres

Paramètre	Description	Valeur par défaut
Support minimal	Il vous permet d'indiquer combien de fois une règle doit être trouvée dans le jeu de données pour être considérée comme une recommandation valable.	2
Confiance minimale	Il vous permet d'indiquer la confiance en-dessous de laquelle une règle ne sera pas utilisée comme recommandation. Par défaut, une règle sera utilisée comme recommandation uniquement si le produit B a été acheté avec le produit A plus de la moitié du nombre de fois où le produit A a été acheté.	0,5

Paramètre	Description	Valeur par défaut
KI minimal	La capacité prédictive (KI) vous permet d'indiquer un seuil de qualité en-dessous duquel une règle ne sera pas considérée comme une recommandation valable.	0

5.11.1.2 Récapitulatif des paramètres de modélisation

Cet écran résume les paramètres de modélisation.

Nom du modèle	Le nom du modèle est renseigné automatiquement. Il correspond au nom de la variable cible, suivi d'un trait de soulignement ("_") et du nom de la source de données, sans son extension.
Description	Une fois le modèle enregistré, la description vous permettra de reconnaître votre modèle lorsque vous voudrez le rouvrir à l'aide de SAP Predictive Analytics.
Modélisation	Ce champ non-modifiable indique quel type de moteur sera utilisé pour créer le modèle.
Sauvegarde automatique...	Ce bouton ouvre le panneau *Sauvegarde automatique* qui vous permet d'indiquer que le modèle doit être enregistré automatiquement à la fin du processus d'apprentissage et de définir tous les paramètres requis. Lorsque la sauvegarde automatique est activée, une coche verte s'affiche sur le bouton *Activer la sauvegarde automatique*. À la fin du processus d'apprentissage, un message d'avertissement s'affiche, confirmant que le modèle a bien été enregistré.
Exporter le script KxShell...	Ce bouton ouvre le panneau vous permettant de générer un script KxShell qui reproduit le modèle en cours.

5.11.2 Utilisation du modèle

5.11.2.1 Moteur de recommandation

Cet écran affiche les différentes options proposées pour visualiser, appliquer, enregistrer ou exporter le modèle.

Affichage

Aperçu du modèle	affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.
Rapports statistiques	En plus des rapports statistiques détaillés, vous obtenez un ensemble de tables vous permettant d'analyser plus finement votre modèle.
Paramètres du modèle	Toutes les informations nécessaires au codage des variables, la compression, le remplacement des variables manquantes, la fréquence des variables, les paramètres du modèle, etc., sont stockées dans un arbre des paramètres.
Visualisation des nœuds	Les recommandations reposent sur une analyse des réseaux sociaux. Cette option vous permet d'afficher le graphe social pour un nœud spécifique et de naviguer à travers ce graphe.
Recommandation	permet à l'utilisateur de sélectionner un client du jeu de données et de visualiser les recommandations correspondantes.

Exécution

Application du modèle	un nouveau jeu de données peut être scoré en utilisant le modèle créé. Les scores peuvent également être appliqués aux données utilisées pour créer le modèle.
Génération du code	Le codage des variables et les paramètres du modèle peuvent être stockés en tant que code qui peut être compilé et utilisé sur d'autres systèmes pour modéliser des données sans utiliser SAP Predictive Analytics.
Exporter le script KxShell	Génère un script KxShell qui peut être utilisé pour recréer le modèle et l'appliquer.
Enregistrer le modèle	Un nouveau jeu de données peut être scoré en utilisant le modèle créé. Les scores peuvent également être appliqués aux données utilisées pour créer le modèle.

5.11.2.2 Affichage de nœuds pour Recommendation

Les recommandations fournies par Recommendation reposent sur des graphes d'analyse des réseaux sociaux. Ce panneau vous permet d'afficher le graphe concernant un produit donné.

- Affichage d'un graphe [page 283]
- Enregistrement d'un graphe [page 283]
- Développement des nœuds [page 284]
- Personnalisation de l'affichage des nœuds [page 284]

- Affichage de graphes [page 284]
- Comparaison de deux graphes [page 284]

5.11.2.2.1 Affichage d'un graphe

- Pour afficher un graphe :
  1. Entrez ou sélectionnez l'identifiant du nœud dans le champ *Nœud* situé dans l'onglet *Naviguer* du ruban.
     Remarque : Le menu déroulant liste par ordre décroissant les 10 nœuds ayant le plus haut degré de connectivité dans le graphe choisi.
  2. Dans le cas d'un graphe *transactionnel*, sélectionnez la colonne où se trouve l'identifiant dans la liste *Ensemble de nœuds*.
  3. Sélectionnez le graphe que vous voulez afficher dans la liste *Graphe*. Par défaut, la valeur est *Tous les graphes*.
  4. Cliquez sur *Afficher le nœud* et sélectionnez l'option désirée. Les graphes correspondants sont affichés dans différents onglets dans la partie droite du panneau.
  5. Cliquez deux fois sur un nœud pour afficher ses voisins.

> **i** Remarque
>
> En cliquant une fois sur un nœud, vous pouvez voir le nombre de ses voisins sur le côté gauche du panneau sans les afficher. Cela peut être utile pour des nœuds ayant un grand nombre de voisins.

- Pour afficher plus de détails :
  1. Dans l'onglet *Reporting* du ruban, cliquez sur *Rapports*.
  - Sélectionnez les informations que vous souhaitez afficher.

5.11.2.2.2 Enregistrement d'un graphe

- Pour exporter un graphe dans un fichier CSV :
  - Dans l'onglet *Reporting* du ruban, cliquez sur le bouton *CSV* pour exporter le graphe affiché dans un fichier CSV. Deux fichiers seront créés : le premier décrit les liens et le second décrit les nœuds.
- Pour copier un graphe :
  - Dans l'onglet *Reporting* du ruban, cliquez sur *Copier* pour copier le graphe affiché dans le presse-papier.
- Pour enregistrer un graphe dans une image :
  - Dans l'onglet *Reporting* du ruban, cliquez sur le bouton *Image* pour enregistrer le graphe affiché dans un fichier PNG.

5.11.2.2.3 Développement des nœuds

- Pour développer tous les nœuds :
  1. Faites un clic droit sur l'arrière-plan du graphe pour afficher le menu contextuel.
  2. Sélectionnez l'option *Développer les nœuds* > *Développer tous les nœuds*.
- Pour développer uniquement les nœuds sélectionnés :
  1. Sur le graphe, sélectionnez les nœuds que vous souhaitez développer en utilisant la touche Maj de votre clavier.
  2. Faites un clic droit sur l'arrière-plan du graphe pour afficher le menu contextuel.
  3. Sélectionnez l'option *Développer les nœuds* > *Développer les nœuds sélectionnés*.
- Pour annuler une action :
  1. Sur le graphe, sélectionnez les nœuds que vous souhaitez développer en utilisant la touche Maj de votre clavier.
  2. Sélectionnez l'une des options suivantes :
     - *Replier* : vous pouvez également utiliser le raccourci clavier `Ctrl` + `Z`
     - *Redéployer* : vous pouvez également utiliser le raccourci clavier `Ctrl` + `Y`

5.11.2.2.4 Personnalisation de l'affichage des nœuds

- Pour personnaliser l'affichage des nœuds :
  - Dans l'onglet *Affichage* du ruban, cliquez sur *Format des nœuds*. La fenêtre *Paramètres d'affichage* s'ouvre.

5.11.2.2.5 Affichage de graphes

Vous pouvez afficher plusieurs graphes en même temps en n'utilisant qu'une partie du panneau pour chaque graphe.

- Pour afficher plusieurs graphes :
  - Cliquez sur une des icônes d'onglet et bougez-la pour pouvoir visualiser plus d'un graphe ou filtre à la fois.

5.11.2.2.6 Comparaison de deux graphes

Recommendation offre la possibilité de comparer deux graphes et de visualiser la différence entre les deux. Cette fonctionnalité peut être utile, par exemple, pour connaître l'évolution des nœuds, si les graphes ont un filtre sur la variable date.

- Pour comparer deux graphes :
  - Cliquez sur le bouton pour comparer deux graphes. La couleur donne des informations supplémentaires sur les liens :
    - *vert* : le lien existe uniquement dans le premier graphe
    - *rouge* : le lien existe uniquement dans le second graphe
    - *noir* : le lien existe dans les deux graphes et le poids est affiché de la façon suivante : `<weight in graph1> => <weight in graph2>`

5.11.2.3 Recommandation

Générer les recommandations

- Pour générer les recommandations pour un client donné
  1. Dans la liste *Recommandeur*, sélectionnez le recommandeur que vous souhaitez utiliser pour générer les recommandations.
  2. Dans le champ portant le nom de la variable contenant les identifiants des clients, saisissez l'identifiant du client pour lequel vous voulez afficher les recommandations. L'identifiant doit exister dans le jeu de données d'entraînement.
  3. Les options décrites ci-dessous vous permettent d'adapter la liste des recommandations à vos besoins :
     - *Ne pas recommander si déjà possédé*
       Par défaut les produits déjà achetés par le client ne seront pas recommandés.

     Décochez cette option si vous voulez inclure ces produits dans les recommandations.
     - *Inclure les meilleures ventes* :
       Par défaut, les meilleures ventes ne sont pas incluses dans les recommandations. La recommandation des produits les mieux vendus n'est pas liée aux clients mais provient d'une analyse spécifique.

     Cochez cette option si vous souhaitez que les meilleures ventes soient incluses dans les produits recommandés.
     - *Garder les N plus importants*
       Par défaut, seules les cinq meilleures recommandations sont fournies dans le jeu de données de résultats.
       - Pour modifier le nombre de recommandations fournies dans les résultats, changez le nombre indiqué dans le champ *Garder les N plus importantes*.
  4. Pour obtenir la liste complète des recommandations générées, décochez cette option. Attention, il peut y avoir un grand nombre de recommandations pour un client, il faut donc être prudent lorsque vous désactivez cette option. Cliquez sur *Accéder aux recommandations*. La liste des recommandations correspondant aux options que vous avez sélectionnées est affichée en bas de l'écran.

Comprendre la liste des recommandations

La colonne...	contient...
Recommandation	les produits recommandés pour le client sélectionné. Par défaut, les recommandations sont triées par ordre décroissant de confiance.
Confiance	la probabilité que l'utilisateur achète un produit/service en fonction de son panier.

5.11.2.4 Application d'un modèle

Ce panneau permet d'appliquer le modèle à un nouveau jeu de données et de générer des recommandations pour vos clients d'après l'historique de leurs achats.

- Paramétrage des options de génération [page 286]
- Sélection du jeu de données d'application [page 287]
- Format du jeu de données d'entrée [page 288]
- Jeu de données d'entrée contenant de nouveaux clients [page 287]
- Enregistrement des résultats générés par le modèle [page 288]
-
- Table des résultats [page 290]

5.11.2.4.1 Paramétrage des options de génération

- Pour paramétrer les options de génération :
  1. Dans la liste *Recommandeur*, sélectionnez le recommandeur que vous souhaitez utiliser pour générer les recommandations.
  2. Paramétrer les options suivantes pour adapter la liste des recommandations à vos besoins :
     - *Ne pas recommander si déjà possédé*
       Par défaut les produits déjà achetés par le client ne seront pas recommandés.

     Décochez cette option si vous voulez inclure ces produits dans les recommandations.
     - *Inclure les meilleures ventes* :
       Par défaut, les meilleures ventes ne sont pas incluses dans les recommandations. La recommandation des produits les mieux vendus n'est pas liée aux clients mais provient d'une analyse spécifique.

     Cochez cette option si vous souhaitez que les meilleures ventes soient incluses dans les produits recommandés.
     - *Garder les N plus importants*
       Par défaut, seules les cinq meilleures recommandations sont fournies dans le jeu de données de résultats.
       - Pour modifier le nombre de recommandations fournies dans les résultats, changez le nombre indiqué dans le champ *Garder les N plus importantes*.

Pour obtenir la liste complète des recommandations générées, décochez cette option. Attention, il peut y avoir un grand nombre de recommandations pour un client, il faut donc être prudent lorsque vous désactivez cette option.

5.11.2.4.2 Sélection du jeu de données d'application

- Pour sélectionner le jeu de données d'application
  1. Dans la section *Jeu de données d'application*, cliquez sur le bouton *Parcourir* sur la ligne *Dossier* pour sélectionner le sous-répertoire du fichier source.
  2. Cliquez sur le bouton *Parcourir* sur la ligne *Données* pour sélectionner le fichier texte ou la source ODBC. Le format des données d'entrée doit être identique à celui utilisé pour créer le modèle.

> **i** Remarque
>
> Un message d'avertissement peut s'afficher indiquant que le jeu de données contient plus de champs que déclaré dans le modèle. Il s'agit d'un simple message d'information et ne requiert aucune action de votre part. Cliquez sur **Fermer** pour poursuivre.

5.11.2.4.3 Jeu de données d'entrée contenant de nouveaux clients

Dans le cas d'un nouveau client, vous ne pouvez pas personnaliser les recommandations. Vous devez générer un ensemble de règles correspondant à chaque produit. À partir du moment où un nouveau client clique sur un produit, l'ensemble de règles correspondant est appliqué. Bien que le client n'ait pas encore acheté le produit, en cliquant dessus il indique qu'il est intéressé.

Voici un exemple d'ensemble de règles :

PRODUIT_A	PRODUIT_B
Bière	Frites
Bière	Couches
...	...

- Pour générer un ensemble de règles :
  1. Reprenez vos données de transaction. Le tableau ci-dessous est un exemple de jeu de données transactionnelles qui relie les clients et les produits.

TRANSACTION_ID	CLIENT_ID	PRODUIT_ID
1	Julien	Bière
2	Julien	Cigarettes

TRANSACTION_ID	CLIENT_ID	PRODUIT_ID
3	Marco	Couches
4	Marco	Bière
5	Jim	Cigarettes
6	Jim	Bière
...	...	...

2. Pour chaque produit, conservez une seule référence. Dans l'exemple ci-dessus, on obtient :

CLIENT_ID	PRODUIT_ID
Julien	Bière
Julien	Cigarettes
Marco	Couches
...	...

3. Appliquez votre modèle Recommandation sur le fichier dédoublonné. L'ensemble de règles est généré.

> **i** Remarque
>
> Vous pouvez générer un ensemble de règles avec SQL si vous travaillez avec une base de données, ou avec un fichier Microsoft Excel si vous travaillez avec un fichier plat.

5.11.2.4.4 Format du jeu de données d'entrée

Dans **Recommandation**, un jeu de données d'application d'entrée doit contenir deux colonnes :

- Une première colonne contenant les identifiants des clients pour qui l'on souhaite générer des recommandations. Vous devez dédoublonner la colonne pour que chaque client n'apparaisse qu'une seule fois dans le jeu de données.
- Une deuxième colonne qui ne devrait pas être renseignée mais qui doit avoir un en-tête (colonne_2 par exemple).

> **i** Remarque
>
> Il est à noter que la colonne *KxComIndex* n'est pas nécessaire pour appliquer un modèle *Recommandation*.

5.11.2.4.5 Enregistrement des résultats générés par le modèle

- Pour enregistrer les résultats générés par le modèle :

- Dans la section *Résultats générés par le modèle*, sélectionnez le type du jeu de données de sortie dans la liste *Type de données*.
- Cliquez sur le bouton *Parcourir* sur la ligne *Dossier* pour spécifier le répertoire qui contiendra les résultats.
- Saisissez un nom pour le fichier de sortie dans le champ *Données*, ou utilisez le bouton *Parcourir* pour sélectionner un fichier.
- Cliquez sur *Appliquer*.

> **i** Remarque
>
> Un message d'avertissement peut s'afficher indiquant que le jeu de données contient plus de champs que déclaré dans le modèle. Il s'agit d'un simple message d'information et ne requiert aucune action de votre part. Cliquez sur *Fermer* pour poursuivre.

5.11.2.4.6 Mode Application in-database

Ce mode optimisé permet d'appliquer le modèle directement dans la base de données. Cela évite d'avoir à extraire les données de la base de données et accélère le processus d'écriture des données de sortie du modèle.

Le mode Application in-database est disponible pour :

- Modèles de classification/régression avec un ordre du polynôme égal à 1.
- Modèles de segmentation générés avec des expressions SQL
- Recommandation et modèles sociaux

Vous pouvez utiliser ce mode si :

- Le jeu de données d'application d'entrée (table, vue, requête, manipulation de données) et le jeu de données de résultats (tables) provient de la même base de données.
- Au moins une variable du jeu de données d'entraînement utilisé pour créer le modèle est déclarée comme clé dans Automated Analytics.

  > **i** Remarque
  >
  > La (Les) clé (s) définie(s) dans le modèle est (sont) identifiée(s) en tant que clé primaire dans la table de sortie.
- L'utilisateur Automated Analytics qui déclenche la tâche d'application dispose des droits de lecture de l'entrée d'application et d'écriture de sortie d'application dans la base de données.
- Il existe un générateur d'indicateurs pour la base de données.
- Recommandation et modèles sociaux uniquement : Le modèle doit être enregistré dans la base de données contenant les jeux de données d'application.

  > **i** Remarque
  >
  > Si le modèle n'a pas été enregistré, une boîte de dialogue s'affiche et vous propose de l'enregistrer. Si vous n'enregistrez pas le modèle, Automated Analytics passe automatiquement au processus d'application standard.

> **i** Remarque
>
> Lorsque vous cochez l'option *Utiliser l'application directe dans la base de données*, l'option *Ajouter un écart de score* est automatiquement sélectionnée. Vous pouvez ainsi vérifier, pour chaque variable et chaque catégorie de variable, les écarts entre le modèle et le jeu de données d'entrée utilisé pour l'application du modèle.

5.11.2.4.7 Table des résultats

Chaque ligne du jeu de données de sortie correspond à une recommandation pour un client donné. Les deux premières colonnes du jeu de données de résultats proviennent du jeu de données d'application :

- La première colonne contient l'identifiant du client pour qui est faite la recommandation.
- La deuxième colonne contient l'identifiant du produit dont l'achat par le client a permis de lui faire cette recommandation.

Les autres colonnes du jeu de données de résultats sont générées automatiquement :

Colonne	Description
`sn_rec_rule_id`	L'identifiant de la règle qui a été utilisée pour recommander ce produit
`sn_rec_kxReco`	Le **produit recommandé** pour le client
`sn_rec_source`	Le recommandeur qui a été utilisé pour générer la recommandation
`sn_rec_score`	Le score de la recommandation, c'est-à-dire la confiance de la règle utilisée pour générer cette recommandation

5.11.2.5 Génération du code

Cet écran vous permet de générer un fichier de code compilable à partir du modèle créé. Le modèle créé sera identique au modèle généré par SAP Predictive Analytics.

Paramètres standard

Le fichier de code généré par SAP Predictive Analytics contiendra toutes les informations nécessaires pour le modèle, comme le codage des variables, les valeurs de remplacement des valeurs manquantes, les compressions et les paramètres du modèle.

- Pour générer le code correspondant au modèle
  1. Sélectionnez dans la liste *Recommandeur* le recommandeur à appliquer.
  2. Vous pouvez personnaliser la liste des recommandations en définissant les options suivantes :
     - *Ne pas recommander si déjà possédé*
       Par défaut les produits déjà achetés par le client ne seront pas recommandés. Décochez cette option si vous voulez inclure ces produits dans les recommandations.

- *Inclure les meilleures ventes*
  Par défaut, les meilleures ventes ne sont pas incluses dans les recommandations. La recommandation des produits les mieux vendus n'est pas liée aux clients mais provient d'une analyse spécifique. Cochez cette option si vous souhaitez que les meilleures ventes soient incluses dans les produits recommandés.
- *Garder les N plus importants*
  Par défaut, seules les cinq meilleures recommandations sont fournies dans le jeu de données de résultats. Pour modifier le nombre de recommandations fournies dans les résultats, changez le nombre indiqué dans le champ *Garder les N plus importantes*.

3. Pour obtenir la liste complète des recommandations générées, décochez cette option. Le nombre de recommandations pouvant être élevé, nous vous recommandons de faire attention lorsque vous désactivez cette option. Dans la section *Paramètres du code*, sélectionnez le type de code à générer.
4. Dans la section *Sortie*, cliquez sur le bouton *Parcourir* associé au champ *Dossier* et sélectionnez le dossier où vous voulez enregistrer le fichier généré.
5. Dans le champ *Fichier généré*, saisissez le nom du fichier exporté. Pour remplacer un fichier existant, sélectionnez-le à l'aide du bouton *Parcourir*.
6. Si vous avez sélectionné l'option *Visualiser le code généré*, celui-ci s'affiche à la fin de la génération.

5.11.2.6 Chargement d'un recommandeur

Ce panneau vous permet de recharger des recommandeurs déjà créés.

- Pour ouvrir un recommandeur
  1. Dans la liste *Type de données*, sélectionnez le type de stockage du modèle.
  2. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où le modèle est enregistré. Dans le cas d'une base de données protégée, saisissez les informations d'identification dans les champs *Utilisateur* et *Mot de passe* affichés dans la fenêtre *Sélection d'une source de données*. Tous les recommandeurs enregistrés dans cette source seront affichés.

     > **i** Remarque
     >
     > Lorsque vous affichez le contenu d'une base de données, vous pouvez être amené à cliquer sur le bouton **Actualiser** pour voir tous les modèles, en particulier les plus récents.

  3. Sélectionnez le recommandeur à charger.
  4. Cliquez sur le bouton *Ouvrir*. L'écran *Moteur de recommandation* s'affiche. Toutes les options pour afficher les résultats du recommandeur sélectionné sont disponibles.
- Pour supprimer des recommandeurs
  1. Sélectionnez le recommandeur que vous voulez supprimer (utilisez la touche CTRL du clavier pour sélectionner plusieurs modèles).
  2. Cliquez sur le bouton *Supprimer*.

5.12 Calcul

Cet écran s'affiche lorsque vous vous reconnectez au serveur pendant la génération ou l'application d'un modèle. Le message affiché dépend de l'action effectuée par le serveur.

6 Expert Analytics

6.1 Aide en ligne d'Expert Analytics

Concernant une documentation détaillée sur Expert Analytics, reortez-vous à .

7 Toolkit

7.1 Explorateur de jeux de données

L'explorateur de jeux de données permet de visualiser les jeux de données existants et de les supprimer si nécessaire.

- Pour afficher la liste des jeux de données
  1. Dans la liste déroulante, sélectionnez le type de jeux de données que vous souhaitez parcourir (fichier texte, bases de données, etc.).
  2. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de données contenant les jeux de données.
     La liste des jeux de données s'affiche dans le champ *Jeux de données*.

Visualisation d'un jeu de données

- Pour visualiser un jeu de données
  1. Cliquez sur le jeu de données que vous souhaitez visualiser. Le bouton *Aperçu* en bas du panneau devient actif.
  2. Cliquez sur le bouton *Voir les données*. Les 100 premières lignes du jeu de données s'affichent dans une nouvelle fenêtre.
- Pour sélectionner le nombre de lignes à afficher
  1. Dans le champ *Première ligne*, saisissez le numéro de la ligne précédant la première ligne que vous souhaitez afficher (par exemple, pour commencer à la ligne 1, entrez 0).
  2. Dans le champ *Dernière ligne*, saisissez le numéro de la ligne précédant la dernière ligne que vous souhaitez afficher (par exemple, pour terminer à la ligne 100, entrez 99).
  3. Cliquez sur le bouton *Actualiser* pour afficher les lignes sélectionnées.
- Pour chercher une variable
  1. Cliquez sur le bouton *Rechercher* (symbolisé par une loupe). La fenêtre *Rechercher une variable par* s'affiche.
  2. Sélectionnez le type de recherche que vous voulez effectuer (par index ou par nom). L'index d'une variable est son numéro de colonne dans le jeu de données.
  3. Sélectionnez dans la liste déroulante proposée l'index ou le nom de la variable.

Cliquez sur le bouton *OK*. La colonne de la variable recherchée apparaît sélectionnée dans la fenêtre *Aperçu des données*.

- Pour afficher les statistiques des variables des premières lignes du jeu de données
  1. Dans le champ *Nombre de lignes*, sélectionnez le nombre de ligne pour lesquelles vous voulez voir les statistiques. Pour afficher les statistiques de l'ensemble du jeu de données, saisissez un chiffre inférieur à 1.

2. Cliquez sur le bouton *Statistiques* (icône graphe). La fenêtre *Statistiques pour*<DataSetName>*sur les premières*<n>*lignes* s'ouvre.

- Suppression d'un jeu de données
  1. Cliquez sur le jeu de données à supprimer Le bouton *Supprimer* en bas du panneau devient actif.
  2. Cliquez sur le bouton *Supprimer*. Une boîte de dialogue s'affiche vous demandant de confirmer la suppression.
  3. Cliquez sur le bouton *Oui*. Le jeu de données est supprimé.

7.2 Effectuer un transfert de données

7.2.1 Sélectionner la source du jeu de données

Ce panneau permet de sélectionner le jeu de données que vous voulez répliquer dans un format différent. Vous pouvez dupliquer une partie seulement du jeu de données d'origine en utilisant l'échantillonnage.

- Pour sélectionner un jeu de données
  1. Sélectionnez le type de fichier de jeu de données dans la liste *Type de données*.
  2. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de données contenant le jeu de données.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Données* pour sélectionner le fichier ou la table contenant le jeu de données.
  4. Cliquez sur le bouton *Suivant*.
     Si la source de données que vous êtes en train de transférer est un jeu de données d'analyse, un message s'affiche pour vous demander si vous voulez transférer les données contenues dans le jeu de données d'analyse ou uniquement leur définition. Dans le dernier cas de figure, notez que les tables et les champs requis par le jeu de données d'analyse doivent être accessibles lorsque vous utilisez la définition transférée du jeu de données d'analyse.
- Options avancées
  - Cliquez sur le bouton pour modifier les paramètres avancés (échantillonnage, paramètres des fichiers, définition de la ligne d'en-tête).
  - Cliquez sur le bouton pour afficher le jeu de données ou les statistiques du jeu de données.
  - Cliquez sur le bouton *Définir les correspondances* pour associer les variables trouvées dans le nouveau jeu de données à celles du jeu de données d'origine.

7.2.2 Description du jeu de données

Les données à transférer sont décrites dans ce panneau.

- Définitions [page 296]
- Description des données [page 298]

- Enregistrement d'une description [page 299]
- Structure [page 300]
- Ajout d'un filtre au jeu de données [page 304]
- Traduction des catégories [page 304]

7.2.2.1 Définitions

Description d'une variable

La description précise d'une variable permet aux fonctionnalités d'Automated Analytics d'effectuer les calculs appropriés lors de la génération d'un modèle en vue d'obtenir les meilleurs résultats.

Chaque variable ou attribut est décrit par les informations suivantes :

Champ	Description	Valeurs
Nom	nom de la variable (ne peut être modifié)	
Type	type de valeurs enregistrées dans cette variable Remarque : Lorsque vous déclarez une variable comme date ou date et heure, codeur de date extrait automatiquement de cette variable les informations de date telles que le jour du mois, l'année, le trimestre, etc. Des variables supplémentaires contenant ces informations sont créées lors de la génération et utilisées comme variables d'entrée du modèle. Le codeur de date est désactivé pour les modèles de série temporelle.	*Nombre* : nombres "calculables" (attention : les numéros de téléphone ou de compte ne doivent pas être considérés comme des nombres). *Entier* : nombres entiers *Chaîne* : chaînes de caractères *Date et heure* : date et heure *Date* : dates

Champ	Description	Valeurs
Valeur	décrit le type de valeur Attention : lorsque vous créez un modèle d'encodage textuel, il faut que le jeu de données contienne au moins une variable textuelle pour pouvoir passer au panneau suivant.	*Continue* : variable numérique à partir de laquelle une moyenne, une variance, etc. peuvent être calculées *Nominale* : variable de catégorie (seule valeur possible pour une chaîne) *Ordinale* : variable numérique discrète dont l'ordre relatif est important *Textuelle* : variable textuelle contenant des expressions, des phrases ou des textes complets
Clé	variable clé ou identifiant pour l'enregistrement Il est possible de déclarer des clés multiples qui seront construites selon l'ordre indiqué (1-2-3-...).	0 : la variable n'est pas un identifiant, 1 : identifiant primaire, 2 : identifiant secondaire, ...
Ordre	indique si cette variable représente un ordre naturel. Si la valeur est 1, la variable est utilisée dans les expressions SQL dans une condition "ORDER BY". Avertissement : définir une variable comme variable d'ordre aura une incidence sur le choix d'une stratégie de partition. Seules les stratégies séquentielles peuvent être utilisées avec une variable d'ordre.	0 : la variable ne représente pas un ordre naturel 1 : la variable représente un ordre naturel
Inconnu	Chaîne utilisée dans les données pour représenter une valeur manquante	Par exemple : 999 ou #Vide
Groupe	indique le nom du groupe auquel appartient cette variable. Les variables d'un même groupe contiennent une même information et donc ne seront pas croisées pour les modèles d'un degré supérieur à 1. Ce paramètre sera utilisable dans une version future.	
Description	étiquette de description supplémentaire pour la variable	

Bibliothèque de variables

La bibliothèque de variables vous permet de stocker les descriptions de variables que vous avez déjà utilisées. Vous pourrez ainsi les réutiliser automatiquement lors de la description d'un jeu de données par analyse.

7.2.2.2 Description des données

Il existe deux façons de décrire les données :

- en analysant le jeu de données ;
- en se servant d'une description du jeu de données complet préalablement enregistré sous forme de fichier ou de table de base de données.

Analyse du jeu de données

L'analyse du jeu de données d'entraînement vous permettra de commencer avec une description obtenue soit par l'analyse des cents premières lignes de données, soit par l'utilisation d'une description issue de la bibliothèque de variables si une variable de même nom s'y trouve. Vous pouvez ensuite modifier les descriptions proposées (en vérifiant surtout le genre des variables).

> **i** Remarque
>
> Pour les fichiers texte, il est bon d'avoir une première ligne d'en-tête contenant les noms des variables.

Lorsque vous créez la description par *analyse* automatique, vérifiez chaque variable afin de valider la description comme indiqué dans la définition Description d'une variable.

- Pour analyser le jeu de données
  - Cliquez sur le bouton *Analyser* situé dans l'onglet *Accueil* du ruban.
    Les variables dont la description provient de la bibliothèque de variables sont signalées par l'icône dans la colonne *Index*.

Chargement d'une description du jeu de données

Pour décrire le jeu de données, vous pouvez utiliser un fichier ou une table de base de données défini(e) précédemment et contenant les descriptions de toutes les variables.

- Pour ouvrir une description de jeu de données
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Ouvrir une description*.
  2. Dans la fenêtre *Ouvrir une description pour...*, sélectionnez le format de la description dans la liste déroulante *Type de données*.
  3. Utilisez le bouton *Parcourir* à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où doit être stockée la description.
  4. Utilisez le bouton *Parcourir* situé à droite du champ *Description* pour sélectionner le fichier ou la table contenant la description.
  5. Cliquez sur *OK*. La fenêtre se ferme et la description s'affiche dans le panneau principal. Vous pouvez alors la modifier si nécessaire.

7.2.2.3 Enregistrement d'une description

Vous pouvez enregistrer la description des variables soit dans la bibliothèque de variables, soit dans un fichier ou une table de base de données.

Dans la bibliothèque de variables

Vous pouvez choisir d'enregistrer les descriptions de toutes les variables, ou sélectionner uniquement celles qui vous intéressent. Vous pouvez aussi supprimer des variables de la bibliothèque de variables.

Si vous avez défini un référentiel de métadonnées unique (option accessible via le bouton *Aide*, puis *Modifier les options...*, onglet *Référentiel de métadonnées*), les descriptions sont enregistrées dans ce référentiel sinon elles sont enregistrées au même emplacement que le jeu de données que vous décrivez.

- Pour enregistrer toutes les descriptions :
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Enregistrer dans la bibliothèque de variables*.
  2. Sélectionnez l'option *Toutes les variables*.
- Pour enregistrer uniquement certaines descriptions :
  1. Sélectionnez la variable dont vous voulez enregistrer la description. Pour sélectionner plusieurs variables, utilisez la touche `Ctrl`.
  2. Cliquez sur le bouton *Enregistrer dans la bibliothèque de variables*.
  3. Sélectionnez l'option *Variables sélectionnées*.
- Pour supprimer de la bibliothèque de variables :
  - Sélectionnez la variable à supprimer de la bibliothèque de variables et cliquez sur Supprimer de la bibliothèque de variables.

Dans un fichier ou une table de base de données

Vous pouvez enregistrer l'ensemble des descriptions du jeu de données dans un fichier ou dans une table de base de données.

- Pour enregistrer la description du jeu de données dans un fichier ou une table :
  1. Dans l'onglet *Accueil* du ruban, cliquez sur le bouton *Enregistrer la description*. La fenêtre *Enregistrer une description de...* s'affiche.
  2. Dans la liste *Type de données*, sélectionnez le format dans lequel vous voulez enregistrer la description.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour choisir le dossier ou la base de données où vous souhaitez enregistrer la description.
  4. Dans le champ *Description*, saisissez le nom du fichier ou de la table où vous souhaitez enregistrer la description.
  5. Cliquez sur *OK*.

7.2.2.4 Structure

Une fois la description des données affichée dans le panneau, l'onglet *Structures* est ajouté au ruban, ce qui vous permet de modifier la structure des données des variables.

État

La dernière colonne du tableau de la description indique l'état de la structure de chaque variable. Le tableau suivant énumère les états possibles d'une structure de variable.

Icône	État	Description
	indéfini	Le codage des données va déterminer automatiquement le regroupement de catégories en fonction de leur interaction avec la variable cible.
	non modifiable	La structure d'une variable de chaîne ordinale ne peut être modifiée.
	définie par extraction des statistiques de la variable.	L'utilisateur doit ouvrir et valider la structure.
	définie par l'utilisateur ou importée d'un modèle existant.	

Définir une structure

Il existe plusieurs façons de définir la structure d'une variable :

- en extrayant automatiquement les catégories à partir des statistiques de la variable, puis en modifiant ou en validant la structure extraite,
- en dupliquant la structure d'une autre variable sur la variable sélectionnée,
- en important automatiquement la structure à partir d'un modèle existant contenant les mêmes données.
- en créant une structure à partir de zéro.

L'option *Activer l'optimisation des regroupements basés sur la variable cible effectués par K2C* permet au codage des données de regrouper les groupes de catégories définies dans la structure de variable s'ils apportent les mêmes informations.

> **i** Remarque
>
> La traduction des catégories d'une variable n'a pas d'influence sur sa structure qui doit être définie en fonction des valeurs initiales de la variable.

Extraire la structure d'une variable à partir des statistiques

- Pour extraire la structure d'une variable...

- avec l'utilisation de l'onglet *Structures* du ruban :
  1. Sélectionnez les variables dont vous souhaitez extraire la structure.
  2. Cliquez sur le bouton *À partir des statistiques*. Une barre de progression est affichée au cours de l'extraction de la structure. Une fois l'extraction terminée, les icônes correspondant aux variables concernées changent, indiquant le succès de l'opération et vous permettant de les identifier clairement. Vous pouvez alors modifier la structure selon vos spécifications.
- avec l'utilisation du menu contextuel :
  1. Sélectionnez les variables dont vous souhaitez extraire la structure. Cliquez dans le tableau avec le bouton droit de la souris.
  2. Un menu contextuel s'affiche.
  3. Sélectionnez l'option *Définir une structure* > *À partir des statistiques*. Une barre de progression est affichée au cours de l'extraction de la structure. Une fois l'extraction terminée, les icônes correspondant aux variables concernées changent, indiquant le succès de l'opération et vous permettant de les identifier clairement. Vous pouvez alors modifier la structure selon vos spécifications.

Importer la structure d'une variable à partir d'un modèle existant

- Pour importer la structure d'une variable à partir d'un modèle...
  - avec l'utilisation de l'onglet *Structures* du ruban :
    1. Sélectionnez les variables dont vous souhaitez extraire la structure.
    2. Cliquez sur *À partir d'un modèle* et sélectionnez soit *Variables sélectionnées*, soit *Toutes les variables* selon les variables que vous avez sélectionnées. Le panneau *Chargement d'un modèle* s'affiche.
    3. Dans la liste *Type de données*, sélectionnez le type de stockage du modèle.
    4. Utilisez le bouton *Parcourir* situé à côté du champ *Dossier* pour sélectionner le dossier ou la base de données où est stocké le modèle.
    5. Dans la liste des modèles qui s'affiche, sélectionnez le modèle que vous voulez utiliser pour importer les structures de variables.
    6. Cliquez sur le bouton *OK*.
    7. Dans la liste *Cible du modèle chargé*, sélectionnez la cible du modèle. Les variables que vous avez sélectionnées sont affichées dans une liste avec les variables correspondantes du modèle chargé. Vous pouvez ajouter ou supprimer des variables de cette liste et visualiser la structure des variables du modèle, comme décrit ci-dessous.
    8. Une fois que toutes les variables pour lesquelles vous souhaitez importer la structure sont affichées dans la liste, cliquez sur le bouton *OK*. La fenêtre de sélection se ferme et l'icône de structure est mise à jour.
  - avec l'utilisation du menu contextuel :
    1. Faites un clic droit sur les variables dont vous souhaitez extraire la structure. Un menu contextuel s'affiche.
    2. Cliquez sur *À partir d'un modèle* et sélectionnez soit *Variables sélectionnées*, soit *Toutes les variables* selon les variables que vous avez sélectionnées. Le panneau *Chargement d'un modèle* s'affiche.
    3. Référez-vous aux instructions listées au-dessus pour continuer.
- Pour ajouter une variable à la liste
  1. Dans la liste *Variable du modèle*, sélectionnez la variable que vous voulez ajouter à la liste des variables pour lesquelles la structure sera importée du modèle.
  2. Cliquez sur le bouton *Ajouter*. La variable s'affiche dans la liste.

- Pour supprimer une variable de la liste
  1. Dans la liste des variables pour lesquelles la structure sera importée du modèle, sélectionnez la variable que vous voulez supprimer.
  2. Cliquez sur le bouton *Supprimer*. La variable est supprimée de la liste et ajoutée à la liste *Variable du modèle chargé*.
- Pour visualiser la structure d'une variable définie par le modèle chargé
  - Si la variable n'a pas encore été ajoutée à la liste des variables pour lesquelles la structure sera importée du modèle :
    1. Dans la liste *Variable du modèle*, sélectionnez la variable pour laquelle vous souhaitez importer la structure.
    2. Cliquez sur le bouton *Afficher*, la structure de la variable s'ouvre dans une nouvelle fenêtre.
  - Si la variable est déjà dans la liste des variables pour lesquelles la structure sera importée du modèle :
    1. Cliquez deux fois sur la variable dont vous souhaitez voir la structure.

Créer une structure

- Pour créer ou modifier la structure d'une variable...
  - à l'aide de l'onglet *Structures* du ruban :
  - Cliquez sur *Nouvelle Structure*. La fenêtre de modification s'ouvre.
  - Utilisation de la colonne *Structure*
  - Cliquez deux fois sur l'icône Structure correspondant à la variable dont vous voulez modifier ou créer la structure. La fenêtre de modification s'ouvre.
    Si vous aviez extrait la structure des statistiques de la variable ou d'un modèle, les champs de la fenêtre sont déjà préremplis.

Structures par type de variable

Étant donné que la structure d'une variable dépend de son type de valeur (nominale, continue, ordinale), l'éditeur de structure diffère pour chaque type.

Structure d'une variable continue

La structure d'une variable continue est définie par plusieurs intervalles composés chacun de :

- une borne inférieure (*[ ]*) ouverte ou fermée,
- une valeur minimale (*Minimum*),
- une valeur maximale (*Maximum*)
- une borne supérieure (*[ ]*) ouverte ou fermée.

Tous les intervalles doivent être adjacents : il ne doit y avoir aucun trou ni chevauchement entre deux intervalles.

L'option *Ajouter manquante* vous permet d'indiquer avec quel intervalle les valeurs manquantes devront être regroupées.

L'option *Inclure les données plus petites*, vous permet d'inclure dans le premier intervalle les données plus petites que celui-ci. De la même façon l'option *Inclure les données plus grandes*, vous permet d'inclure dans le dernier intervalle les données plus grandes que celui-ci.

- Pour créer un intervalle
  1. Cliquez sur le bouton *Ajouter* pour créer un intervalle. La fenêtre de modification s'ouvre.
  2. Sélectionnez le type de la borne de début en cliquant sur le bouton *]*.
  3. Saisissez la valeur minimale de l'intervalle dans le champ de texte de gauche.
  4. Saisissez la valeur maximale de l'intervalle dans le champ de texte de droite.
  5. Sélectionnez le type de la borne de début en cliquant sur le bouton *[*.
  6. Cochez la case *Manquante* si les valeurs manquantes doivent être regroupées avec cet intervalle.
  7. Cliquez sur le bouton *Oui* pour valider votre intervalle.
- Pour séparer un intervalle
  1. Sélectionnez un intervalle.
  2. Cliquez sur le bouton *Séparer*. L'intervalle sélectionné est automatiquement séparé en deux intervalles égaux.
- Pour fusionner plusieurs intervalles
  1. Sélectionnez les intervalles à fusionner. Ceux-ci doivent être adjacents.
  2. Cliquez sur le bouton *Joindre*.
- Pour supprimer un intervalle
  1. Sélectionnez un ou plusieurs intervalles. Ceux-ci doivent être adjacents.
  2. Cliquez sur le bouton *Supprimer*. Les intervalles précédent et suivant sont étendus pour inclure les valeurs contenues dans les intervalles supprimés.

Structure d'une variable ordinale

La structure d'une variable ordinale est similaire à celle d'une variable continue, à l'exception des limites qui sont toujours fermées et ne peuvent pas être modifiées.

> **i** Remarque
>
> La structure d'une variable de chaîne ordinale ne peut être modifiée.

Structure d'une variable textuelle

La structure des variables textuelles ne peut être modifiée.

Structure d'une variable nominale

La structure d'une variable nominale est composée de plusieurs groupes rassemblant les catégories de la variable.

- Pour créer un groupe de catégories
  1. Dans la section *Structure du groupe*, saisissez dans le champ de texte le nom du groupe à créer. Utilisez la touche `Ctrl` de votre clavier pour sélectionner plusieurs catégories.
  2. Cliquez sur le bouton *Ajouter un nouveau groupe*. Un groupe contenant les catégories sélectionnées est créé dans la liste *Structure du groupe*.
- Pour inclure les valeurs manquantes dans un groupe
  1. Dans la liste *Structure du groupe*, sélectionnez le groupe auquel vous voulez ajouter les valeurs manquantes.
  2. Cliquez sur le bouton *Ajouter manquante*, situé sous la liste *Édition des catégories*. La catégorie *KxMissing*, qui représente les valeurs manquantes, est ajoutée au groupe sélectionné et le bouton *Ajouter manquante* est désactivé. Comme toute catégorie, la catégorie *KxMissing* peut appartenir à un seul groupe à la fois.

- Pour créer une catégorie et l'ajouter à un groupe
  1. Dans la liste *Structure du groupe*, sélectionnez le groupe auquel vous voulez ajouter la nouvelle catégorie.
  2. Dans le champ à droite du bouton *Nouvelle catégorie*, saisissez le nom de la catégorie à ajouter.
  3. Cliquez sur le bouton *Nouvelle catégorie*. La nouvelle catégorie est créée et assignée au groupe sélectionné.
- Pour supprimer un groupe
  1. Dans la liste *Structure du groupe*, sélectionnez le groupe à supprimer.
  2. Cliquez sur le bouton *Supprimer le groupe*. Toutes les catégories appartenant à ce groupe sont rajoutées à la liste *Édition des catégories*, sauf celles qui ont été créées par l'utilisateur.
- Pour supprimer une catégorie d'un groupe
  1. Dans la liste *Structure du groupe*, sélectionnez les catégories à supprimer du groupe.
  2. Cliquez sur le bouton *Supprimer la catégorie*. Toutes les catégories assignées à ce groupe réapparaissent dans la liste *Édition des catégories*, à l'exception de celles créées par l'utilisateur.

> **i** Remarque
>
> - Pour permettre au codage des données (K2C) de réunir les groupes de catégories définis dans la structure des variables s'ils apportent les mêmes informations, cochez le mode *Activer l'optimisation des regroupements basés sur la variable cible effectués par K2C*.
> - Lorsque le mode *Utilise le codage naturel* est activé, seule la version d'origine de la variable est utilisée. Les versions codées sont désactivées et les critères d'exclusion sont relâchés pour les versions d'origine.

7.2.2.5 Ajout d'un filtre au jeu de données

Si vous ne souhaitez pas utiliser le jeu de données entier, vous pouvez créer un filtre pour ne retenir que les données souhaitées.

- Ajouter un filtre
  1. Cochez l'option *Ajouter un filtre au jeu de données*.
  2. Cliquez sur le bouton *Suivant*.

7.2.2.6 Traduction des catégories

Vous pouvez traduire les catégories des variables nominales, enregistrer la traduction ou charger une traduction existante. Cette traduction n'influence pas la structure de la variable, qui doit être définie en fonction des valeurs originales de la variable.

> **i** Remarque
>
> La variable "Clé cible", utilisée par exemple dans les paramètres avancés, ne prend pas en compte une éventuelle traduction quand les valeurs possibles de cette variable sont affichées. Pour cette raison des

valeurs entrées manuellement ne peuvent pas être traitées correctement, si elles ne correspondent pas aux valeurs d'origine.

- Traduire les catégories de variables
  1. Sélectionnez la variable nominale dont vous souhaitez traduire les catégories. Vous pouvez :
     - Soit aller sur l'onglet Édition et cliquer sur *Traduire les catégories*.
     - Soit effectuer un clic droit pour afficher un menu contextuel, puis cliquer sur *Traduire les catégories*.
       Le panneau *Traduire les catégories de <nom_de_la_variable>* s'affiche.
  2. Cliquez sur le bouton *Ajouter une langue* pour sélectionner les langues que vous voulez traduire. Les langues ajoutées s'affichent en colonnes.

     > **i Remarque**
     >
     > Par défaut, la langue de l'interface utilisateur est affichée sous forme de colonne.

  3. Cliquez sur le bouton pour extraire les catégories de variables du jeu de données.
  4. Traduisez les catégories.

     > **i Remarque**
     >
     > - Vous n'avez pas besoin de renseigner tous les champs.
     > - Vous pouvez ajouter ou supprimer des catégories avec les boutons *Ajouter une catégorie* et *Supprimer la catégorie*.

  5. Cliquez sur le bouton *OK*.
- Pour enregistrer la traduction des catégories
  1. Traduisez les catégories de variables comme expliqué ci-dessus.
  2. Cliquez sur le bouton *Enregistrer*.
  3. Choisissez un *Type de données*.
  4. Sélectionnez un *Dossier*.
  5. Entrez un *Nom* pour le fichier ou la table.
  6. Cliquez sur le bouton *OK*.
- Pour charger un fichier de traduction existant
  1. Sélectionnez la variable nominale à traduire. Vous pouvez :
     1. Soit aller à l'onglet Édition et cliquer sur le bouton *Traduire les catégories*.
     2. Soit effectuer un clic droit pour afficher un menu contextuel. Cliquez ensuite sur *Traduire les catégories*.
        Le panneau *Traduire les catégories de <nom_de_la_variable>* s'affiche.
  2. Cliquez sur le bouton *Charger*.
  3. Sélectionnez le format de la traduction dans la liste *Type de données*.
  4. Utilisez le bouton *Parcourir* à droite du champ *Dossier* pour sélectionner le dossier ou la base de données où doit être stockée la description.
  5. Utilisez le bouton *Parcourir* à droite du champ *Table ou fichier* pour sélectionner le fichier ou la table contenant la description.
  6. Cliquez sur le bouton **OK**.
  7. Cliquez sur le bouton *Mettre à jour* pour actualiser l'affichage des catégories.

8. Si les colonnes ne sont pas nommées correctement, utilisez les Paramètres avancés pour choisir la ligne d'en-tête et actualisez à nouveau.
9. Mettez les noms des langues en correspondance avec les langues de la traduction chargée en cliquant sur les catégories et en choisissant la langue qui correspond dans le menu contextuel.
10. Cliquez sur le bouton *OK*.

Paramètres avancés

- Pour définir une ligne d'en-tête
  1. Cliquez sur l'onglet *Ligne d'en-tête*.
  2. Cochez l'option *Forcer la ligne d'en-tête*.
  3. Dans le champ *Ligne*, saisissez le numéro de la ligne que vous voulez utiliser comme ligne d'en-tête.
  4. Cliquez sur *OK*.

7.2.3 Création d'une copie de jeu de données

Ce panneau vous permet de dupliquer un jeu de données dans un fichier ou une table de base de données.

- Pour enregistrer le jeu de données
  1. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de donnée qui contiendra les informations enregistrées.
  2. Saisissez le nom du fichier à créer ou utilisez le bouton *Parcourir* à droite du champ *Jeu de données de sortie* pour sélectionner un fichier existant.
  3. Cliquez sur le bouton Suivant.
     Une boîte de message contenant une barre de progression apparaît. Une fois le fichier ou la table créé, le menu principal s'affiche.

7.2.4 Transfert de données

Ce panneau vous permet de suivre le processus de transfert des données à l'aide d'une barre de progression ou d'un journal. La barre de progression s'affiche par défaut. Pour voir le journal, cliquez sur le bouton *Affiche le détail du processus* situé dans la barre d'outils sous le titre du panneau.

Une fois le transfert de données terminé, cliquez sur le bouton *Suivant* pour retourner au menu principal.

7.2.5 Transfert de manipulation de données

Ce panneau vous permet lors du transfert de la définition d'un jeu de données d'analyse de choisir la base de donnée cible, définir l'association entre les table sources et les tables cibles, et, si nécessaire, l'association entre les champs des tables.

- Pour définir un transfert
  1. Pour sélectionner la base de données cible (c'est-à-dire celle dans laquelle vous voulez transférer la définition du jeu de données d'analyse), cliquez sur le bouton *Parcourir*, puis sélectionnez une base de données dans la boîte de dialogue. Cliquez sur le bouton *OK* pour valider votre sélection.
  2. Cliquez sur le bouton Remplir situé en bas du panneau. L'assistant de modélisation essaie de trouver dans la base de données cible les tables utilisées dans la définition du jeu de données d'analyse en se basant sur leur nom.
  3. Pour chaque table non trouvée ou incorrectement identifiée, cliquez sur la cellule correspondante dans la colonne *Table cible*. Un sélecteur de table est affiché.
  4. Lorsque toutes les associations sont définies, cliquez sur *Vérifier*. L'assistant de modélisation vérifie alors que les associations sont correctes. Si une erreur est détectée, un message de diagnostic est affiché, sinon le bouton *Suivant* est activé.
  5. Cliquez sur le bouton *Suivant* pour effectuer le transfert.

Correspondance des champs

Lorsque les champs de la table cible ne portent pas le même nom que les champs de la table source, il est nécessaire de définir l'association entre les champs de la table source et les champs de la table cible.

La colonne *Définir la correspondance des champs* vous indique l'état de l'association des champs.

- : la définition du transfert n'a pas encore été validée ou l'association des champs par défaut est correcte.
- : l'association des champs a été modifiée et est correcte.
- : l'association des champs n'est pas correcte (qu'elle ait été modifiée ou non).
- Pour définir l'association des champs
  1. Dans la colonne *Définir la correspondance des champs* cliquez sur l'icône . Une fenêtre permettant de modifier la correspondance des champs s'ouvre alors.
  2. Cliquez sur le bouton Remplir. L'assistant de modélisation va alors essayer d'associer les champs en se basant sur la ressemblance des noms et sur le format de stockage des champs.
     Remarque : Deux champs doivent avoir le même format de stockage pour pouvoir être associés.
  3. Pour changer les associations qui vous semblent fausses ou qui n'ont pas pu être trouvées :
     1. Cliquez dans la colonne *Champ cible* à la ligne voulue.
     2. Dans la liste déroulante, sélectionnez le champ cible que vous voulez associer au champ source.
  4. Lorsque toutes les associations sont définies, cliquez sur *Valider*. Si toutes les correspondances sont correctes, le bouton *OK* est activé, sinon un message de diagnostic est affiché.
  5. Cliquez sur le bouton *OK* pour valider l'association des champs.

7.2.6 Sélection des variables

Cet écran permet de choisir les variables de votre jeu de données que vous voulez conserver ou exclure.

- Pour exclure une variable
  1. Sélectionnez la variable à exclure dans la liste *Variables explicatives conservées*.
  2. Cliquez sur le bouton **>** situé à gauche de la section *Variables exclues*.

> **i** Remarque
>
> Toutes les listes peuvent être triées par ordre d'apparition dans le fichier ou par ordre alphabétique.

7.3 Obtenir des statistiques descriptives pour un jeu de données

7.3.1 Chargement du jeu de données d'entrée

La fonctionnalité *Statistiques Descriptives* vous permet de générer les statistiques des variables d'un jeu de données, telles que les statistiques croisées ou les courbes de profit lorsqu'une variable est utilisée pour estimer la variable cible. Ce panneau permet de sélectionner le jeu de données à analyser.

- Pour sélectionner un jeu de données
  1. Sélectionnez le type de fichier de jeu de données dans la liste *Type de données*.
  2. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de données contenant le jeu de données.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Données* pour sélectionner le fichier ou la table contenant le jeu de données.
  4. Cliquez sur le bouton *Suivant*.
     Une confirmation peut vous être demandée avant l'exécution de la requête si vous utilisez une base de données Teradata, Oracle ou SQLServer 2005. Cela est dû au *mode Explain* qui estime la charge de tous les requêtes SQL avant de les exécuter. Pour en savoir plus sur cette fonctionnalité, voir le Mode Explain.

     > **i** Remarque
     >
     > Pour modifier les paramètres du mode Explain, contactez votre administrateur SGBD.
- Options avancées
  - Cliquez sur le bouton pour modifier les paramètres avancés (échantillonnage, paramètres des fichiers, définition de la ligne d'en-tête).
  - Cliquez sur le bouton pour afficher le jeu de données ou les statistiques du jeu de données.
  - Cliquez sur le bouton *Définir les correspondances* pour associer les variables trouvées dans le nouveau jeu de données à celles du jeu de données d'origine.

7.3.2 Paramètres avancés du modèle

Ce panneau vous permet de sélectionner une variable à utiliser comme estimateur de variable cible ou de spécifier des index de clusters.

Estimateur de cible

Un estimateur de cible est une variable permettant de prédire le comportement de la variable cible. Utiliser un tel estimateur équivaut à créer un modèle.

S'il n'existe aucune variable pouvant être utilisée comme estimateur dans le jeu de données, ne remplissez rien dans ce panneau.

Remarque : Seules des variables continues peuvent être utilisées comme estimateurs.

- Pour sélectionner un estimateur de la cible
  1. Sélectionnez l'onglet *Estimateur de cible*.
  2. Dans la colonne *Variables*, cliquez sur la variable à utiliser comme estimateur.
  3. Cliquez dans la case correspondante de la colonne *est estimateur de*. Une liste des variables cible existantes s'affiche.
  4. Sélectionnez la variable cible à estimer.

Estimateur de cluster

Lorsque vous utilisez un jeu de données qui résulte de l'application d'un modèle de segmentation et dans lequel chaque enregistrement est assigné à un cluster, vous pouvez utiliser cet onglet pour spécifier quelle colonne du jeu de données contient les identifiants de clusters. Cela peut être utilisé pour générer des statistiques telles que la taille des clusters et la somme des carrés résiduels pour ces clusters. Un des usages les plus évidents est la comparaison des résultats de mise en cluster obtenus grâce à différents outils ou différentes méthodes de calcul.

- Pour ajouter un estimateur de cluster
  1. Sélectionnez l'onglet *Estimateur de cluster*.
  2. Cliquez sur le bouton *Ajouter un estimateur*. Une variable est ajoutée à la liste des estimateurs.
  3. Cliquez sur la variable pour afficher la liste de toutes les variables nominales du jeu de données.
  4. Sélectionnez dans la liste la variable contenant un identifiant de cluster.
  5. Si le modèle de mise en cluster utilisé pour générer l'identifiant de cluster était un modèle de segmentation (donc avec une variable cible), sélectionnez la variable cible dans la liste *Cible du cluster*.
  6. Répétez les étapes 2 à 5 si votre jeu de données contient plus d'un identifiant de cluster. Par exemple, si le jeu de données contient les résultats de deux outils de mise en cluster différents, vous pouvez comparer directement les statistiques sur le même jeu de données.
  7. Dans la liste *Distance* sous la liste des estimateurs, sélectionnez la distance utilisée pour évaluer la proximité de deux individus donnés dans le modèle de mise en cluster original.

8. Dans la liste *Codage* sous la liste des estimateurs, sélectionnez le codage utilisé dans le modèle de mise en cluster original. Le codage par défaut utilisé lors de la génération d'un modèle de mise en cluster avec Automated Analytics est :
   - *Moyenne de la cible* pour les modèles de segmentation,
   - *Regroupement* pour les autres modèles.

   > **i** Remarque
   >
   > Si le jeu de données contient des résultats de plusieurs modèles, ceux-ci doivent avoir été générés avec les mêmes métriques de distance et le même codage.
9. Une fois tous les estimateurs indiqués, cliquez sur le bouton *Valider*.

- Pour supprimer un estimateur de cluster
  1. Sélectionnez la variable d'estimation dans la liste.
  2. Cliquez sur le bouton *Supprimer un estimateur*.

7.3.3 Analyse des statistiques

Cet écran vous permet d'afficher les statistiques générées pour le jeu de données, de vérifier les écarts dans un nouveau jeu de données et d'enregistrer les statistiques.

Affichage

- *Aperçu du modèle* : affiche les informations du modèle généré, telles que le nombre de variables, les indicateurs de performance, les informations cible, etc.

  > **i** Remarque
  >
  > Si vous n'avez pas utilisé d'estimateur de la cible, les indicateurs de performance ne sont pas renseignés.
- *Courbes de performance* : selon le type de cible, le graphique des courbes de performances vous permet de :
  - Visualiser le profit réalisable par rapport à votre problématique en utilisant le modèle généré lorsque la cible est nominale.
  - Comparer les performances du modèle généré à celles d'un modèle aléatoire et d'un modèle parfait hypothétique quand la variable cible est nominale.
  - Comparer la valeur prédite à la valeur réelle dans le cas d'une variable cible continue.

    > **i** Remarque
    >
    > Les courbes de performance ne sont pas proposées si vous n'avez pas sélectionné un estimateur de la cible auparavant.
- *Importance des catégories* : ce panneau détaille l'importance de chaque catégorie pour les variables nominales ou ordinales et l'importance de chaque classe pour les variables continues.

- *Tables d'analyse* : SAP Predictive Analytics vous fournit un ensemble de tables vous permettant d'analyser plus finement votre jeu de données.
- *Paramètres du modèle* : toutes les informations nécessaires au codage des variables, la compression, le remplacement des variables manquantes, la fréquence des variables, les paramètres du modèle, etc., sont stockées dans un arbre des paramètres.
- *Matrice de confusion* : vous permet de comparer les valeurs cible prédites par le modèle avec les valeurs réelles et de définir le score au-dessus duquel les observations seront considérées comme positives : la valeur cible correspond à celle souhaitée. Elle vous permet également de faire des simulations de profit selon le score choisi comme seuil ou d'adapter automatiquement le seuil pour obtenir un profit maximal.
- *Exploration des règles d'association* : permet d'effectuer des recherches parmi toutes les règles d'association trouvées en utilisant des options de filtrage.
- *Rapports statistiques* : permet de visualiser des tableaux d'analyse détaillés du modèle selon le type de modèle que vous avez généré.
- *Tuiles* : toutes les informations nécessaires à la géolocalisation.

Exécution

- *Vérification des écarts* : l'utilisateur peut charger un nouveau jeu de données et vérifier s'il existe des écarts par rapport au jeu de données ayant été utilisé pour créer le modèle.

Enregistrer

- Les statistiques peuvent être enregistrées et chargées ultérieurement comme n'importe quel modèle SAP Predictive Analytics.

7.3.4 Tuiles

Le panneau **Tuiles** vous permet de diviser l'espace en tuiles.

Les options disponibles dans le panneau sont indiquées dans le menu du ruban, lequel est divisé en 4 onglets : **Accueil**, **Export**, **SIG** et **WFS**.

- Pour sélectionner les tuiles à afficher :
  1. Dans la liste **Variable de position**, sélectionnez la variable composite que vous avez définie dans le panneau **Description de données**. La variable de position présente à la fois une longitude et une latitude.
  2. Dans la liste **Jeu de données**, sélectionnez le jeu de données pour lequel vous souhaitez visualiser les vignettes.
  3. Dans la liste **Cible**, sélectionnez la cible de votre choix. Si vous souhaitez obtenir des statistiques descriptives pour un jeu de données et non des groupes optimisés pour une cible, alors il n'y a pas de cible.

4. En fonction de vos données, le nombre de tuiles peut être très élevé, rendant le résultat affiché difficile à interpréter. Si vous ne souhaitez pas visualiser toutes les tuiles en détails, cochez la case **Grouper les tuiles**.
   Notez que les tuiles sont regroupées en fonction de la cible.

Le tableau suivant vous donne un aperçu des titres des colonnes et ce à quoi ils correspondent.

Titre de la colonne	Description
Nom de tuile	Les tuiles sont numérotées dans l'ordre dans lequel le chemin a été parcouru.
Description	Correspond à la latitude et à la longitude constituant une tuile
Effectif pondéré	Nombre de lignes correspondant à un point situé dans une tuile
Fréquence de tuile	Pourcentage de lignes dans une tuile
Densité	Rapport de la surface de la tuile sur l'effectif pondéré
Moyenne de la cible	Moyenne des cibles positives dans une tuile

Cochez la case *Ne pas utiliser la notation scientifique* si vous voulez que les valeurs affichées dans le tableau soient arrondies.

- Accueil [page 312]
- Export [page 313]
- SIG [page 314]
- WFS [page 314]

7.3.4.1 Accueil

- Pour sélectionner des tuiles :
  1. Dans l'onglet *Accueil*, cliquez sur *Sélection*.
  2. Sélectionnez l'un des options suivantes :
     - *Sélectionner/Désélectionner*
     - *Sélectionner tout*
     - *Tout désélectionner*
- Pour copier le tableau ou le graphe dans le presse-papier :
  1. Dans l'onglet Accueil, cliquez sur *Copier*.
  2. Sélectionnez l'option de votre choix.
- Pour faire un aperçu des tuiles :
  1. Sélectionnez les tuiles que vous voulez visualiser.
  2. Cliquez sur *Aperçu de la sélection*.
     Une fenêtre s'ouvre dans Google Earth. Les tuiles sont affichées avec leur latitude et longitude.
- Pour changer de vue :

À noter que le panneau *Tuiles* affiche le tableau par défaut.

1. Cliquez sur *Changer de vue* pour afficher le diagramme à barres. Notez que le diagramme à barre compare les tuiles entre elles sur la base d'une valeur qui peut être définie.
2. Dans la liste *Colonne affichée*, sélectionnez la valeur qui sera utilisée pour la comparaison avec les tuiles.

- Pour afficher les métriques :
  - Cliquez sur *Métriques*. Une fenêtre affichant les informations sur les tuiles s'ouvre. Elle indique la variable de position, le jeu de données et la cible que vous avez sélectionnés ainsi que la moyenne et la médiane des tuiles.

7.3.4.2 Export

Pour chaque tuile, vous pouvez définir la couleur et l'opacité des tuiles en fonction des métriques des tuiles.

Par exemple, vous pouvez sélectionner *Cibles positives* dans la liste *Couleur*, et *Fréquence* dans la liste *Opacité*.

- *Couleur*
  Dans le cas d'un nombre élevé de cibles positives dans une tuile, la tuile sera *rouge*. Inversement, dans le cas d'un nombre restreint de cibles positives dans une tuile, la tuile affichera une couleur *verte*.
- *Opacité*
  De même, plus la fréquence est élevée, plus la tuile sera *opaque.*

- Pour définir la couleur :
  - Dans l'onglet *Export* du ruban, sélectionnez la couleur dans la liste correspondante.
- Pour définir l'opacité :
  - Dans l'onglet *Export*, sélectionnez l'opacité dans la liste correspondante.
- Pour exporter les tuiles :
  1. Dans la liste déroulante *Type de fichier*, sélectionnez le type de fichier vers lequel vous voulez exporter les données des tuiles.
     Le tableau suivant récapitule les types de fichier possibles :

Type de fichier	Description
KML	KML (Keyhole Markup Language) est un langage basé sur le formalisme XML et destiné à la gestion de l'affichage de données géospatiales dans les logiciels de SIG.
GML	Le Geography Markup Language (GML) est un langage dérivé du XML pour encoder, manipuler et échanger des données géographiques.
SHAPE_FILE	Shape File, ou "fichier de formes", est un format de données vectorielles géospatiales très répandu pour les systèmes d'informations géographiques.

  2. Cliquez sur *Export*.
  3. Sélectionnez l'option désirée selon ce que vous souhaitez exporter. Une fenêtre *Enregistrer* s'ouvre.
  4. Choisissez le dossier où le fichier sera exporté.
  5. Cliquez sur *OK*.

7.3.4.3 SIG

Un SIG (système d'information géographique ou GIS en anglais) est un système qui permet de capturer, d'organiser, d'analyser et d'afficher tous types de données géographiques.

- Pour ajouter des tuiles :
  1. Dans l'onglet SIG, cliquez sur *Ajout de tuiles*.
  2. Sélectionnez l'option de votre choix.
- Pour ajouter une icône :

> **i** Remarque
>
> Pour utiliser le SIG, vous devez définir une adresse dans le champ *Protocole du système d'information géographique* dans les options de l'application.

1. Dans le tableau, sélectionnez la tuile de votre choix.
2. Cliquez sur *Ajout d'icône*. Le champ *URI générée* est automatiquement renseigné avec l'adresse que vous avez définie dans les options de l'application (*Préférences*), suivi des coordonnées du centre de gravité de la tuile, c'est-à-dire le point annulant une égalité vectorielle.
3. Cliquez sur *Ouvrir*.

- Pour faire une capture d'écran :
  1. Dans le tableau, sélectionnez la tuile de votre choix.
  2. Cliquez sur le bouton *Capture d'écran*. La fenêtre *Chemin de la capture d'écran* s'ouvre.
  3. Entrez le chemin de votre choix dans le champ *Chemin d'enregistrement de la capture d'écran sur le serveur*.
  4. Cliquez sur *OK*. Le chemin apparaît dans le champ *URI générée*.
  5. Cliquez sur *Ouvrir*.

7.3.4.4 WFS

Vous pouvez paramétrer une adresse de GeoServer et envoyer des tuiles à cette adresse.

Définitions

Terme	Définition
WFS	Le WFS (Web Feature Service) est un protocole décrit par l'Open Geospatial Consortium. Le WFS vous permet de demander des objets géographiques sur le Web via une interface afin de les manipuler ou de récupérer leurs données.

Terme	Définition
GeoServer	GeoServer est un logiciel d'accès libre qui vous permet de partager, de modifier et de publier des données géospatiales. L'application est conçue pour l'interopérabilité avec GeoServer.
Workspace	Un workspace, ou espace de travail, est un conteneur qui permet d'organiser des éléments. Par exemple, un workspace peut être utilisé pour regrouper des layers.
Layer	Les layers ou couches réfèrent à des données matricielles ou vectorielles présentant des objets géographiques. Une layer représente un objet géographique. Chaque layer a une source de données, appelée stockage.

Envoyer des tuiles

Il faut avoir définir un workspace et un layer dans *GeoServer* pour envoyer des tuiles.

- Pour envoyer les tuiles :
  1. Définissez les paramètres *GeoServer*.
  2. Dans l'onglet *WFS*, cliquez sur *Envoyer les tuiles*.
  3. Sélectionnez l'option de votre choix. Un message confirme que la transaction *WFS* a été complétée.

7.4 Répertorier des valeurs distinctes dans un jeu de données

7.4.1 Description de vos données

Les métadonnées peuvent être chargées au format texte délimité par des tabulations ou des virgules, extraites d'une source ODBC, ou encore être dans un format propriétaire. Il existe deux façons de décrire les données :

- *Analyser* : les cent premières lignes de données sont analysées et une première détermination de la description du fichier de données est fournie à l'utilisateur. Pour les fichiers texte, il est bon d'avoir une première ligne d'en-tête contenant les noms des variables. Une fois le fichier de description créé, le bouton *Enregistrer la description* permet de l'enregistrer dans le sous-dossier indiqué dans la case *Dossier*. Le nom du fichier de description peut être entré dans la case *Description*.
- *Ouvrir une description* : permet de charger un fichier de description enregistré auparavant.
- *Aperçu des données* : affiche les cent premières lignes du fichier de données pour vous aider à valider la description.

Chaque variable ou attribut est décrit par les informations suivantes :

- *Nom* : nom de la variable (ne peut être modifié)

- *Stockage* : type de valeurs enregistrées dans cette variable
  - *Nombre* : contient uniquement des nombres "calculables" (attention : les numéros de téléphone ou de compte ne doivent pas être considérés comme des nombres).
  - *Chaîne* : la variable contient des chaînes de caractères.
  - *Date et heure* : la variable contient des dates et heures.
  - *Date* : la variable contient des dates.
- *Valeur* : décrit le type de valeur
  - *Continue* : variable numérique à partir de laquelle une moyenne, une variance, etc. peuvent être calculées
  - *Nominale* : variable de catégorie (seule valeur possible pour une chaîne)
  - *Ordinale* : variable numérique discrète dont l'ordre relatif est important
- *Clé* : variable clé ou identifiant pour l'enregistrement (0 : la variable n'est pas un identifiant, 1 : identifiant primaire, 2 : identifiant secondaire...)
- *Ordre* : indique si cette variable représente un ordre naturel (0 : la variable ne représente pas un ordre naturel ; 1 : la variable représente un ordre naturel). Si la valeur est 1, la variable est utilisée dans les expressions SQL dans une condition "ORDER BY".
  Attention : définir une variable comme variable d'ordre aura une incidence sur le choix d'une stratégie de partition. Seules les stratégies séquentielles peuvent être utilisées avec une variable d'ordre.
- *Manquante* : chaîne utilisée dans les données pour représenter une valeur manquante (par exemple, "999" ou "`#Empty`", à entrer sans les guillemets).
- *Groupe* : indique le nom du groupe auquel appartient cette variable. Les variables d'un même groupe contiennent une même information et donc ne seront pas croisées pour les modèles d'un degré supérieur à 1. Ce paramètre sera utilisable dans une version future.

Description : étiquette de description supplémentaire pour la variable

7.4.2 Définition des données d'entrée

Ce panneau vous permet de sélectionner le jeu de données à utiliser pour extraire des informations sur une variable.

- Pour sélectionner un jeu de données
  1. Sélectionnez le type de fichier de jeu de données dans la liste *Type de données*.
  2. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de données contenant le jeu de données.
  3. Utilisez le bouton *Parcourir* situé à droite du champ *Données* pour sélectionner le fichier ou la table contenant le jeu de données.
  4. Cliquez sur le bouton *Suivant*.

7.4.3 Sortie

Ce panneau permet d'enregistrer les catégories de variable extraites d'un jeu de données.

- Pour enregistrer les catégories de variables
  1. Utilisez le bouton *Parcourir* situé à droite du champ *Dossier* pour sélectionner le dossier ou la base de donnée qui contiendra les informations enregistrées.
  2. Saisissez le nom du fichier à créer ou utilisez le bouton *Parcourir* à droite du champ *Jeu de données de sortie* pour sélectionner un fichier existant.
  3. Cliquez sur le bouton *Suivant*.
     Une boîte de message contenant une barre de progression apparaît. Une fois le fichier ou la table créé, le menu principal s'affiche.

7.4.4 Paramètres

Ce panneau vous permet de choisir la variable dont vous souhaiter extraire les catégories.

- *Pour sélectionner une variable*
  1. Sélectionnez la variable dans la liste *Nom de la colonne de référence*.
  2. Cliquez sur le bouton *Suivant*.

8 Définition des préférences

8.1 Présentation des menus

Les menus supérieurs vous permettent d'effectuer les actions suivantes :

- modifier les options de l'application telles que la langue d'interface, les emplacements de stockage pour le jeu de données d'entrée et de sortie de l'application, le référentiel de métadonnées ainsi que le thème graphique ;
- personnaliser vos feuilles de style ;
- contacter le Support technique en cas de problème ;
- afficher des informations sur SAP Predictive Analytics.

Éléments de menu

Menu	Panneau	Fonctionnalité	Raccourcis clavier
Fichier	*Préférences*	Définit les options de l'application.	F2
	Journal d'erreurs	Affiche la liste des erreurs et le journal.	F4
	Quitter	Ferme l'application	ALT + F4
Aide	*Obtenir de l'aide pour ce panneau*	Affiche l'aide contextuelle pour l'écran d'application en cours.	F1
	Rechercher dans l'aide complète	Ouvre l'aide en ligne sur la plateforme web SAP. Vous pouvez utiliser l'arborescence de navigation ou le moteur de recherche pour trouver les informations dont vous avez besoin.	Sans objet
	Contacter le Support	Affiche des informations relatives au Support technique.	F3
	Saisir la clé de licence	Utilisez cette option pour saisir les clés de licence pour Automated Analytics et Expert Analytics.	Sans objet

Menu	Panneau	Fonctionnalité	Raccourcis clavier
	À propos de	Affiche des informations relatives au produit.	Sans objet

8.2 Préférences

8.2.1 Définition des préférences d'application

Utilisez la commande *Préférences* pour accéder aux panneaux d'options permettant de définir les préférences d'application pour Automated Analytics. La commande *Préférences* est accessible depuis tous les panneaux de l'Assistant de modélisation :

Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.

Cliquez sur l'icône correspondante pour accéder aux options que vous souhaitez modifier.

Vous pouvez également exécuter les actions suivantes :

- Le bouton *Réinitialiser* permet de restaurer les paramètres par défaut du panneau d'options en cours.
- Le bouton *Réinitialiser toutes les options* permet de restaurer les paramètres par défaut de tous les panneaux d'options.

Informations associées

Définition de la délégation à APL pour l'entraînement des modèles [page 323]
Définition et modification d'une structure de variable [page 327]
Personnalisation de la feuille de style des rapports [page 331]
Paramétrage des options de géolocalisation [page 334]
Paramétrage des options générales [page 319]
Définition des options du référentiel de métadonnées [page 325]
Modification du contenu de la bibliothèque de variables [page 325]
Paramétrage des options de graphique [page 330]
Paramétrage des options de stockage [page 322]

8.2.2 Paramétrage des options générales

Utilisez le panneau *Général* pour définir des préférences telles que la langue d'interface, le pays et les niveaux de message.

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Définissez les options à l'aide des valeurs possibles décrites dans la table.

Option	Définition	Valeurs possibles
Pays	Définit le format de dates et de nombres spécifique au pays.	• AT (Autriche) • BR (Brésil) • CN (Chine) • DE (Allemagne) • ES (Espagne) • FR (France) • GB (Grande-Bretagne) • JP (Japon) • MX (Mexique) • PT (Portugal) • RU (Russie) • TW (Taïwan, Chine) • US (USA) (par défaut)
Langue	Définit la langue d'interface.	• de (allemand) • en (anglais) (par défaut) • es (espagnol) • fr (français) • ja (japonais) • pt (portugais) • ru (russe) • zh (chinois)

Option	Définition	Valeurs possibles
Niveau de message	Indique le niveau des messages à afficher par l'interface. Les messages affichés sont ceux qui correspondent à tous les niveaux, de 0 jusqu'au niveau indiqué. • 0 : messages d'erreur affichés dans une fenêtre pop-up • 1 : informations affichées dans une fenêtre pop-up • 2 : obsolète • 3 : message d'information enregistré dans le fichier journal • 4 : informations détaillées enregistrées dans le fichier journal • 5 : messages d'avertissement enregistrés dans le fichier journal • 6 : messages de progression affichés dans une barre de progression **i Remarque** Chaque niveau affiche également les messages de niveau précédent (exemple : le niveau 1 affiche des messages d'information et des messages d'erreur).	• 6 (par défaut) • Tout entier compris entre 0 et 6.
Taille maximale du journal	Le nombre maximal de lignes affichées dans le journal lors de la génération ou l'application d'un modèle. Lorsque le nombre défini est atteint, les plus anciennes lignes sont supprimées.	• 10 000 (par défaut) • Un entier positif
Niveau de message pour les valeurs aberrantes	Définit le niveau des messages à utiliser lorsque des valeurs aberrantes sont détectées dans les jeux de données lors de l'application d'un modèle. Les valeurs aberrantes sont des valeurs manquantes ou des valeurs hors plage dans le jeu de données d'application par rapport au jeu de données d'apprentissage. Lorsque le niveau affiche un nombre supérieur à 6, les messages n'apparaissent pas dans l'interface.	• 12 (par défaut) • Un entier positif
Afficher l'arbre des paramètres	Indique s'il faut ou non afficher l'arbre des paramètres dans le panneau de débriefing du modèle.	Non sélectionné par défaut
Taille maximale de l'historique des derniers répertoires ouverts	Indique le nombre d'emplacements de stockage à afficher dans l'historique.	• 5 (par défaut) • Un entier positif
Toujours quitter sans confirmer	Indique s'il faut ou non confirmer lors de la fermeture de l'application.	Non sélectionné par défaut

Option	Définition	Valeurs possibles
Inclure test dans la stratégie de partition par défaut.	Indique s'il faut ou non utiliser le jeu de données de test dans la stratégie de partition par défaut. La stratégie de partition par défaut standard est "aléatoire sans test". En activant cette option, la stratégie de partition par défaut sera définie sur "aléatoire".	Non sélectionné par défaut
Séparateur d'instructions SQL	Définit le séparateur utilisé dans les instructions SQL générées par l'application.	• Point-virgule (par défaut)

3. Cliquez sur *OK* pour enregistrer vos modifications.

   Lors de la modification des préférences, il est toujours possible de restaurer les paramètres d'origine en cliquant sur le bouton *Réinitialiser*.

8.2.3 Paramétrage des options de stockage

Utilisez le panneau *Emplacements* pour personnaliser les options relatives aux emplacements de stockage (ou sources de données) traités lors de la modélisation.

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Dans le panneau *Emplacements*, modifiez les options selon le tableau suivant :

Option	Définition	Valeurs
Emplacement par défaut du jeu de données d'application d'entrée	Indique l'emplacement à ouvrir par défaut lors de la sélection des jeux de données d'entrée pour une application de modèle.	• *Dernier emplacement utilisé pour le jeu de données d'application* : Permet de définir le dernier emplacement utilisé pour une application de modèle (par défaut). • *Emplacement du jeu de données d'apprentissage du modèle* (valeur par défaut) : Permet de définir l'emplacement contenant le jeu de données d'apprentissage défini en tant qu'emplacement par défaut.
Emplacement par défaut du jeu de données d'application de sortie	Indique l'emplacement à ouvrir par défaut lors de la sélection des jeux de données de sortie pour une application de modèle.	• *Dernier emplacement utilisé pour les données d'application en sortie* : Permet de définir le dernier emplacement utilisé pour une application de modèle (par défaut). • *Emplacement du jeu de données d'apprentissage du modèle* (valeur par défaut) : Permet de définir l'emplacement contenant le jeu de données d'apprentissage défini en tant qu'emplacement par défaut.

Option	Définition	Valeurs
Emplacement par défaut pour enregistrer des modèles	Indique l'emplacement à sélectionner par défaut lors de l'enregistrement d'un modèle.	• *Dernier emplacement utilisé pour enregistrer des modèles* (valeur par défaut) : Permet de définir le dernier emplacement utilisé pour un modèle (par défaut). • *Emplacement du jeu de données d'apprentissage du modèle* : Permet de définir l'emplacement contenant le jeu de données d'apprentissage défini en tant qu'emplacement par défaut.

3. Cliquez sur *OK* pour enregistrer vos modifications.

   Lors de la modification des préférences, il est toujours possible de restaurer les paramètres d'origine en cliquant sur le bouton *Réinitialiser*.

8.2.4 Définition de la délégation à APL pour l'entraînement des modèles

Pour les modèles utilisant SAP HANA en tant que source de données, les calculs d'entraînement des modèles par défaut sont délégués à la bibliothèque SAP Automated Predictive (APL) depuis SAP HANA (lorsque cela est possible).

Les conditions décrites ci-dessous doivent être remplies pour que la délégation d'entraînement puisse avoir lieu. Si la délégation n'est pas possible ou si l'option n'est pas désélectionnée, le moteur de modélisation Automated Analytics effectue les calculs d'entraînement.

Conditions préalables à la délégation APL :

- APL (version 2.4 ou version supérieure) doit être installé sur le serveur de base de données SAP HANA.
- La version minimum requise de SAP HANA est SPS 10 Database Revision 102.02 (SAP HANA 1.00.102.02).
- La version APL doit être identique à celle de la version serveur ou bureau de SAP Predictive Analytics.
- L'utilisateur SAP HANA qui se connecte à la source de données ODBC doit avoir l'autorisation d'exécuter APL.

Pour en savoir plus, voir le Guide *SAP HANA Automated Predictive Library Reference Guide* disponible sur le SAP Help Portal à l'adresse suivante : http://help.sap.com/pa.

> **!** Restriction
>
> Dans certains cas, l'entraînement des modèles n'est pas délégué à APL :
>
> - Dans les modules Recommandation et Analyse de réseaux sociaux.
> - Lorsque le modèle utilise une stratégie de partition personnalisée.
> - Lorsque le modèle utilise l'option de calcul de l'arbre de décision.

Pour désélectionner le comportement par défaut :

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Dans le panneau *Délégation de l'entraînement*, désélectionnez *Remplacer l'apprentissage du modèle par SAP HANA APL si possible*.

3. Cliquez sur *OK* pour enregistrer vos modifications.
   Lors de la modification des préférences, il est possible de restaurer les paramètres d'origine en cliquant sur le bouton *Réinitialiser*.

8.2.5 Configuration de la modélisation native Spark

Grâce à Spark, les modèles Automated Analytics peuvent s'exécuter sous Hadoop avec de meilleures performances, une plus grande évolutivité et pratiquement aucun transfert de données. Pour les modèles utilisant Hadoop en tant que source de données, les calculs d'entraînement des modèles par défaut sont délégués au moteur Spark sous Hadoop (lorsque cela est possible).

Les conditions décrites ci-dessous doivent être remplies pour que la délégation d'entraînement puisse avoir lieu. Si la délégation n'est pas possible ou si l'option n'est pas désélectionnée, le moteur de modélisation Automated Analytics effectue les calculs d'entraînement.

- Vous devez installer SAP Predictive Analytics 2.5 ou une version plus récente.
- Un environnement Big data doit être disponible sur Hadoop.
- Une connexion à Hadoop utilisant les pilotes ODBC Hive doit être configurée selon ce qu'indique le *guide d'accès aux données Predictive Analytics*, section *Connexion à votre système de gestion de bases de données*.
- Dans les versions installées de la distribution Hadoop, Hive et Spark doivent être identiques à ceux mentionnés dans le PAM.
- Les fichiers jar et de connexion "Spark Connectivity" fournis avec l'installation de SAP Predictive Analytics doivent être correctement configurés, tel qu'indiqué dans les guides *Connecting to your Database Management System*

> **i Remarque**
>
> Pour en savoir plus sur la configuration de la connexion à Hive et la modélisation native Spark, voir le *guide d'accès aux données Predictive Analytics*, section *Connexion à votre système de gestion de bases de données*, sur le SAP Help Portal à l'adresse : http://help.sap.com/pa#section5.

> **! Restriction**
>
> Automated Analytics prend uniquement en charge la modélisation native Spark pour les modèles de classification d'entraînement avec cible unique et stratégie de partition aléatoire. Pour tous les autres types de modèles et les fonctionnalités avancées, l'application bascule au processus de modélisation standard.

Pour désélectionner le comportement par défaut :

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Dans le panneau *Délégation de l'entrainement*, désélectionnez *Utiliser modélisation native Spark si possible*.
3. Cliquez sur *OK* pour enregistrer vos modifications.

   Lors de la modification des préférences, il est possible de restaurer les paramètres d'origine en cliquant sur le bouton *Réinitialiser*.

8.2.6 Définition des options du référentiel de métadonnées

Par défaut, les métadonnées (descriptions) associées à un jeu de données sont stockées avec les données. Vous pouvez définir un référentiel distinct dans lequel les descriptions de données sont stockées depuis le panneau d'options *Référentiel de métadonnées*.

Il est également possible de modifier le contenu de la bibliothèque de variables à partir de ce même panneau. Pour en savoir plus, voir la rubrique associée.

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Dans le panneau *Référentiel de métadonnées*, sélectionnez *Mettre les métadonnées dans un endroit spécifique*.
3. Sélectionnez le type d'emplacement de stockage des métadonnées dans la liste déroulante *Type de données*.
4. Dans le champ *Dossier*, accédez à l'emplacement du référentiel de métadonnées.
5. Si l'authentification est requise pour accéder au référentiel de métadonnées, saisissez le nom d'utilisateur et le mot de passe dans les champs *Utilisateur* et *Mot de passe*.

Informations associées

Définition et modification d'une structure de variable [page 327]
Modification du contenu de la bibliothèque de variables [page 325]

8.2.7 Modification du contenu de la bibliothèque de variables

Il est possible de modifier les paramètres de variables et la structure de variable à partir du panneau d'options *Référentiel de métadonnées*.

La bibliothèque de variables vous permet d'enregistrer les descriptions de variables. Lors de la définition d'un jeu de données, si une description existe déjà pour une variable se trouvant dans la bibliothèque, l'application utilise la description. Sinon, la description est estimée à l'aide des cent premières lignes du jeu de données. Une fois le jeu de données décrit, il est possible de stocker une ou toutes les descriptions de variables dans la bibliothèque de variables.

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Dans le panneau *Référentiel de métadonnées*, cliquez sur le bouton *Modifier le contenu de la bibliothèque de variables*.
3. Dans le panneau *Contenu de la bibliothèque de variables*, modifiez les paramètres d'après le tableau ci-dessous. Sélectionnez les valeurs dans la liste déroulante ou cliquez deux fois sur la colonne pour saisir une valeur.

Paramètre	Définition	Valeurs possibles
Stockage	Le type de stockage de la variable.	• nombre • chaîne • entier • date • date/heure • angle
Valeur	Le type de valeur de la variable.	• continue • nominale • ordinale • textuelle
Clé	Le niveau de clé de la variable, où 1 correspond à la clé primaire, 2 à la clé secondaire et ainsi de suite. La valeur zéro (0) signifie que la variable n'est pas une clé.	Entier (commençant à 0)
Ordre	Un nombre différent de 0 indique que la variable peut être considérée comme ordre de tri naturel pour les cas. Exemple de variable : une date ou un ID client. En présence d'une telle variable, les composants SAP Predictive Analytics peuvent utiliser les informations pour améliorer la stratégie de partition.	Entier (commençant à 0)
Manquante	Indique un code spécifique pour les valeurs manquantes. La valeur par défaut généralement utilisée est "null". Cependant, les systèmes de base de données peuvent avoir leur propre code pour les valeurs manquantes, tel que "99" pour les nombres à deux chiffres ou "inconnu" (unknown) pour certaines informations de catégorie.	• chaîne nulle • 99 • inconnu
Groupe	Regroupe les variables appartenant à un seul groupe sémantique. Exemple : en cas de codage disjonctif des variables de catégorie : toutes les variables peuvent se voir affecter des valeurs booléennes, mais uniquement lorsqu'une des colonnes est activée et que toutes les autres sont désactivées. Cela empêche les transformations (et par conséquent les modèles) de prendre en compte des études croisées entre des variables appartenant au même groupe.	
Description	Commentaires décrivant la variable.	

Paramètre	Définition	Valeurs possibles
Structure	Définit le regroupement des catégories de variables. Il est possible d'extraire une structure de variable à l'aide des statistiques des variables ou de définir une structure de variable à partir de zéro. Pour définir une structure de variable, cliquez deux fois sur l'icône de la colonne *Structure*. Pour en savoir plus, voir la rubrique associée.	• Indéfinie Le codage des données Automated Analytics va déterminer automatiquement le regroupement de catégories en fonction de leur interaction avec la variable cible. • Non modifiable La structure d'une variable de chaîne ordinale ne peut être modifiée. • Définie par extraction des statistiques de la variable. • Définie par l'utilisateur.

> **i** Remarque
>
> L'*Index* et le *Nom* des variables stockées dans la bibliothèque ne peuvent pas être modifiés.

4. Cliquez sur *Enregistrer les modifications* pour enregistrer vos modifications.

Informations associées

Définition et modification d'une structure de variable [page 327]

8.2.8 Définition et modification d'une structure de variable

Vous définissez la structure d'une variable en modifiant le contenu de la bibliothèque de variables dans le panneau d'options *Référentiel de métadonnées*.

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Dans le panneau *Référentiel de métadonnées*, cliquez sur le bouton *Modifier le contenu de la bibliothèque de variables*.
3. Sélectionnez la variable dans la table, puis cliquez deux fois sur l'icône de la colonne *Structure*. Une structure de variable peut avoir des états différents, indiqués par les icônes suivantes :
   - Indéfinie Le codage des données Automated Analytics va déterminer automatiquement le regroupement de catégories en fonction de leur interaction avec la variable cible.
   - Non modifiable La structure d'une variable de chaîne ordinale ne peut être modifiée.
   - Définie par extraction des statistiques de la variable.
   - Définie par l'utilisateur.

À propos des types de structures de variables

Valeur de la variable	Structure
Continue	La structure d'une variable continue est définie par plusieurs intervalles composés chacun des éléments suivants : • une borne inférieure ([ ]) ouverte ou fermée, • une valeur minimale (Minimum), • une valeur maximale (Maximum), • une borne supérieure ([ ]) ouverte ou fermée. **i Remarque** Tous les intervalles doivent être adjacents : il ne doit y avoir aucun trou ni chevauchement entre deux intervalles.
Ordinal	La structure d'une variable ordinale est similaire à celle d'une variable continue, à l'exception des limites qui sont toujours fermées et ne peuvent pas être modifiées. **i Remarque** La structure d'une variable de chaîne ordinale ne peut être modifiée.
Nominal	La structure d'une variable nominale est composée de plusieurs groupes rassemblant les catégories de la variable.

Modification de variables continues ou ordinales

La modification d'une variable continue est identique à celle d'une variable ordinale. Une exception cependant : les bornes des variables ordinales sont toujours fermées et ne peuvent pas être modifiées.

Pour...	Procédez ainsi...
Créer un intervalle	1. Cliquez sur le bouton *Ajouter* pour créer un intervalle. 2. La boîte de dialogue *Sélectionner une option* s'affiche. 3. Saisissez la valeur minimale de l'intervalle dans le champ de texte de gauche (]). 4. Saisissez la valeur maximale de l'intervalle dans le champ de texte de droite ([). 5. Cochez la case *Ajouter manquant* si les valeurs manquantes doivent être regroupées avec cet intervalle. 6. Cliquez sur le bouton *Oui* pour valider l'intervalle défini.
Séparer un intervalle	1. Sélectionnez un intervalle. 2. Cliquez sur *Séparer*. L'intervalle sélectionné est automatiquement séparé en deux intervalles égaux.

Pour...	Procédez ainsi...
Fusionner plusieurs intervalles	1. Sélectionnez les intervalles à fusionner. **i Remarque** Ceux-ci doivent être adjacents. 2. Cliquez sur le bouton *Fusionner*.
Supprimer un intervalle	1. Sélectionnez un ou plusieurs intervalles. **i Remarque** Ceux-ci doivent être adjacents. 2. Cliquez sur le bouton *Supprimer*. Pour qu'il n'y ait aucun trou entre eux, les intervalles précédent et suivant sont étendus pour inclure les valeurs contenues dans les intervalles supprimés.

Modification de variables nominales

Pour...	Procédez ainsi...
Créer un groupe	1. Saisissez le nom du groupe dans le champ en regard du bouton *Nouvelle catégorie* (en bas à droite du panneau). 2. Dans la section *Édition des catégories* (en haut à droite du panneau), sélectionnez l'élément créé. 3. Cliquez sur le bouton *Ajouter un nouveau groupe*. Le nouveau groupe s'affiche sous forme de dossier dans la section *Structure du groupe* (côté supérieur gauche) ; il est automatiquement mis en surbrillance.
Inclure les valeurs manquantes dans un groupe	1. Dans la section *Structure du groupe* (côté supérieur gauche), sélectionnez le groupe auquel vous voulez ajouter les valeurs manquantes. 2. Cliquez sur le bouton *Ajouter manquant* (côté inférieur droit du panneau). La catégorie *KxMissing*, qui représente les valeurs manquantes, est ajoutée au groupe sélectionné et le bouton *Ajouter manquant* est désactivé. **i Remarque** La catégorie *KxMissing* peut appartenir à un seul groupe à la fois.
Créer une catégorie et l'ajouter à un groupe	1. Dans la section *Structure du groupe* (côté supérieur gauche), sélectionnez le groupe auquel vous voulez ajouter la nouvelle catégorie. 2. Saisissez le nom de la catégorie dans le champ en regard du bouton *Nouvelle catégorie* (en bas à droite du panneau). 3. Cliquez sur le bouton *Nouvelle catégorie*. 4. Sélectionnez la catégorie nouvellement créée dans la section *Édition des catégories* (en haut à droite du panneau). 5. Cliquez sur le bouton *Ajouter une catégorie*.

Pour...	Procédez ainsi...
Supprimer un groupe	1. Dans la liste *Structure du groupe*, sélectionnez le groupe à supprimer. 2. Cliquez sur le bouton *Supprimer le groupe*. Toutes les catégories appartenant à ce groupe sont de nouveau transférées vers la section *Édition des catégories* (en haut à droite du panneau).
Supprimer une catégorie d'un groupe	1. Dans la section *Structure du groupe* (côté supérieur gauche), sélectionnez la ou les catégories à supprimer du groupe. 2. Cliquez sur le bouton *Supprimer la catégorie*. Les catégories sélectionnées sont supprimées du groupe et sont de nouveau transférées vers la section *Édition des catégories* (en haut à droite du panneau).

8.2.9 Paramétrage des options de graphique

Vous pouvez personnaliser les options d'affichage des graphiques depuis le panneau *Graphique*.

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Dans le panneau *Graphique*, modifiez les options selon la table suivante :

Option	Définition	Valeurs
Thème graphique	Personnalise l'interface utilisateur. Le thème *SAPContrast* est un thème à fort contraste pour les utilisateurs ayant une déficience visuelle. *Aucun* désactive tous les thèmes et vous permet d'utiliser les fonctionnalités d'accessibilité de votre système d'exploitation.	• **SAP** (par défaut) • SAPContrast • Aucun
Taille de la police	Définit la taille de police de l'interface.	• **12** (par défaut) • Entier positif
Nombre de points de la courbe de profits	Nombre de points affichés sur les diagrammes de courbe de profit.	• **20** (par défaut) • Entier positif
Nombre de barres affichées	Nombre de barres affichées par la vue de graphe.	• **1 à 20** (par défaut)
Afficher les diagrammes en 3D	Affiche les diagrammes en 3 dimensions.	• Activé • **Désactivé** (par défaut)
Désactiver le double tampon	Désactive l'optimalisation de l'actualisation de l'interface utilisateur.	• Activé • **Désactivé** (par défaut)
Optimiser pour les affichages distants	Améliore la réactivité de l'interface sur des affichages lents.	• Activé • **Désactivé** (par défaut)

3. Cliquez sur *OK* pour enregistrer vos modifications.

   Lors de la modification des préférences, il est toujours possible de restaurer les paramètres d'origine en cliquant sur le bouton *Réinitialiser*.

8.2.10 Personnalisation de la feuille de style des rapports

Le panneau *Rapport* vous permet de personnaliser l'affichage de vos rapports et des diagrammes en modifiant la feuille de style prédéfinie ou en créant votre propre feuille de style. Il est également possible d'ajouter une bannière de rapports.

Sélectionnez *Fichier* > *Préférences* ou appuyez sur la touche F2, puis sélectionnez le panneau *Rapport*.

Bannière personnalisée

Pour définir une bannière spécialisée pour les rapports, dans le champ *Bannière personnalisée*, sélectionnez le fichier image. Les formats de fichier image disponibles sont les suivants : JPEG, GIF, BMP, PNG et WBMP.

Sélection de la feuille de style active

La feuille de style par défaut est *SAP Predictive Analytics (par défaut)*. D'autres feuilles de styles prédéfinies sont également disponibles. Pour sélectionner la feuille à utiliser, sélectionnez le bouton *Actif* associée à cette feuille de style.

Vous ne pouvez pas modifier les feuilles de styles prédéfinies, mais vous pouvez en faire une copie et l'utiliser en tant que modèle de base pour votre feuille de style personnalisée.

Création et modification d'une feuille de style

1. Avant de pouvoir copier ou créer une feuille de style, vous devez saisir un emplacement pour le référentiel de feuille de style dans *Dossier*.
2. Pour créer une feuille de style, cliquez sur ✚. Une feuille de style nommée `Style Sheet.krs` est ajoutée à la liste des feuilles de style.

   Pour copier une feuille de style, sélectionnez-la dans la liste, puis cliquez sur (Copier). Une feuille de style portant le nom copié et dotée de l'extension `.krs` est ajoutée à la liste des feuilles de style.
3. Sélectionnez la nouvelle feuille de style, puis cliquez sur (Modifier).

4. Dans le panneau *Éditeur de feuilles de style*, personnalisez les principaux aspects visuels du rapport selon le tableau suivant :

Paramètre	Options...	Procédure
Couleur de fond	• choisir une couleur • rendre transparent	Cliquez sur *Opaque* et sélectionnez une couleur dans la zone appropriée ou cliquez sur *Transparent*. **i Remarque** Seuls les formats PDF et HTML peuvent afficher une couleur de fond.
Configuration générale	• Titre • Section • Sous-section • Simple	Cliquez sur le bouton *Modifier* correspondant.
Configuration de tableau	• Taille de la première colonne • Taille du tableau • En-tête du tableau • Nom de ligne de tableau • <number> de ligne	Cliquez sur le bouton *Modifier* correspondant ou saisissez une valeur. Cliquez sur *Ajouter une nouvelle police* pour ajouter une nouvelle police à la ligne.

5. Cliquez sur *Paramètres du diagramme* pour personnaliser les aspects visuels du diagramme selon le tableau suivant :

Pour...	Procédez ainsi...
Ajouter une couleur de diagramme	1. Cliquez sur (Nouveau). Une fenêtre s'ouvre. 2. Configurer la couleur. 3. Cliquez sur *OK*.
Supprimer une couleur de diagramme	1. Sélectionnez la couleur de diagramme. 2. Cliquez sur le bouton (Supprimer).
Modifier une couleur de diagramme	1. Sélectionnez la couleur de diagramme. 2. Cliquez sur (Modifier).
Définir l'orientation de l'histogramme par défaut	Sélectionnez *Horizontale* ou *Verticale* selon vos besoins.

6. Cliquez sur *Sections du rapport* pour personnaliser l'affichage de chaque rapport généré par l'application (*Aperçu*, *Rapport exécutif*, *Analyse*, *Rapport détaillé*, *Aperçu*...).

Pour modifier les sections du rapport, développez une catégorie sur la gauche et sélectionnez un rapport. Le tableau suivant récapitule les différentes options à définir :

Option	Vous permet...	Procédure
Visible	Masquer/afficher les colonnes d'une section de rapport ou même toute une section de rapport. **i Remarque** Au moins une colonne d'une section doit rester visible.	Cochez/décochez la case Visible associée à la section.
Affiché comme	Définit le nom du rapport	Saisissez un titre dans le champ correspondant.
Type de vue	Choisissez la vue Tabulaire, HTML ou Graphique. **i Remarque** L'option Graphique n'est disponible que si la section du rapport peut être affichée comme graphe.	1. Sélectionner une section du rapport. 2. Sélectionnez une option dans la liste déroulante Type de vue.
Type de diagramme	Définir le type de diagramme. **i Remarque** Cette option n'est disponible que pour les sections de type Graphique.	1. Sélectionner une section du rapport. 2. Sélectionnez une option dans la liste déroulante Type de diagramme.
Changer l'orientation des barres	Définit une orientation différente des barres pour une section de rapport spécifique. **i Remarque** Cette option n'est disponible que pour certaines sections de rapport.	Cochez la case Changer l'orientation des barres (si l'option est activée).
Trier par/Ordre de tri	Définir une colonne à trier/classer et choisir l'ordre de tri (croissant ou décroissant).	1. Sélectionner une section du rapport 2. Définit les options de tri dans les listes déroulantes affichées.
Colonnes visibles	Définit les colonnes à afficher dans le rapport.	Sélectionnez les colonnes à afficher.

7. Cliquez sur *Enregistrer* pour enregistrer la feuille de style.
8. Dans le panneau *Rapport*, cliquez sur *OK* pour enregistrer vos modifications.
   Lors de la modification des préférences, il est toujours possible de restaurer les paramètres d'origine en cliquant sur le bouton *Réinitialiser*.

8.2.11 Paramétrage des options de géolocalisation

Utilisez le panneau *Géolocalisation* pour définir l'adresse de base à laquelle envoyer les requêtes GIS (Geographic Information System). Un SIG (ou GIS en anglais) est un système qui requiert une adresse de base et permet de capturer, d'organiser, d'analyser et d'afficher tous types de données géographiques. Un SIG permet par exemple de visualiser des tuiles et séquences sur Google Earth. Une fois l'adresse de base définie, l'URI générée commence automatiquement par cette adresse.

L'adresse définie apparaîtra directement dans les panneaux Tuiles et Visualisation du chemin.

Pour définir une adresse :

1. Sélectionnez *Fichier* > *Préférences*, ou appuyez sur la touche F2.
2. Sélectionnez le panneau *Géolocalisation* et saisissez l'adresse de base dans *Protocole du système d'information géographique*.
3. Cliquez sur *OK* pour enregistrer vos modifications.

8.3 À propos de

Ce panneau contient des informations à propos de SAP Predictive Analytics.

Informations disponibles dans la boîte de dialogue À propos de

Onglet	Fonctionnalités
À propos de	• Nom du produit • Version du produit • Informations sur le produit • Nom du GUI • Version du GUI • Java, machine virtuelle • Modes de modélisation disponibles • Type de licence • Canal • Intermédiaire • Nombre de cœurs achetés • Nombres de cœurs détectés
Licences des générateurs	Générateur de cœur et date d'expiration
Système	Propriété et valeur
Options de lancement	Propriété et valeur
Licences d'accès aux données	Accès aux données et date d'expiration
Licences des composants	Composants et date d'expiration

8.4 Journal d'erreurs

Ce panneau fournit le journal et la liste des erreurs survenues lors de la génération ou l'application d'un modèle. En cas de problème, il vous permet de fournir des informations aux experts du Support SAP.

- Vérification de la liste des erreurs
  1. Cliquez sur l'onglet *Liste des erreurs* pour l'afficher.
  2. Cliquez sur *Copier* pour copier la liste dans le presse-papiers.
  3. Cliquez sur *Vider* pour réinitialiser la liste.
  4. Cliquez sur *Fermer* pour fermer la liste.
- Vérification des journaux
  1. Cliquez sur l'onglet *Journal*. Le journal s'affiche.
  2. Dans la liste déroulante, sélectionnez le modèle pour lequel vous souhaitez afficher le journal.

     Codage des couleurs

Type d'information	Couleur du texte
Informations	noir
Avertissements	orange
Erreurs	rouge

  3. Cliquez sur *Fermer* pour fermer la fenêtre.

> **i** Remarque
>
> Le nombre de lignes affichées dans le journal est limité. Pour augmenter la taille du journal, accédez à *Préférences* > *Général* > *Taille maximale du journal*.

8.5 Contacter le Support

- Pour contacter le Support

Dans l'onglet *Aide* de la barre de menu, cliquez sur le bouton *Contacter le Support...*.

Un panneau de message s'ouvre pour indiquer qu'un rapport de bug a été créé et que vous devez l'envoyer au Support **SAP**.

Clauses de non-responsabilité importantes et informations juridiques

Liens hypertexte

Certains liens affichent une icône et/ou du texte contextuel. Ils fournissent des informations complémentaires.
Explication des icônes :

- Liens accompagnés de l'icône : vous accédez à un site Web non hébergé par SAP. En utilisant de tels liens, vous acceptez (sauf indication contraire expresse dans vos contrats avec SAP) ce qui suit :
  - Le contenu du site vers lequel redirige le lien n'est pas de la documentation SAP. Vous ne pouvez émettre aucune réclamation produit auprès de SAP sur la base de ces informations.
  - SAP n'accepte pas ou désapprouve le contenu affiché sur le site vers lequel vous êtes redirigé, ni ne garantit la disponibilité et l'exactitude dudit contenu. SAP ne saurait être tenue responsable des dommages causés par l'utilisation dudit contenu sauf si de tels dommages étaient causés par une négligence grave ou une faute intentionnelle de SAP.
- Liens accompagnés de l'icône : vous quittez la documentation associée à un produit ou service SAP en particulier et accédez à un site Web hébergé par SAP. En utilisant lesdits liens, vous convenez (sauf indication contraire expresse dans vos contrats avec SAP) que vous ne pourrez pas émettre de réclamation produit auprès de SAP sur la base de ces informations.

Vidéos hébergées sur des plateformes externes

Certaines vidéos peuvent pointer vers des plateformes d'hébergement de vidéos tierces. SAP ne garantit pas la disponibilité future des vidéos stockées sur ces plateformes. Par ailleurs, toute annonce ou tout autre contenu hébergé(e) sur ces plateformes (par exemple, suggestions de vidéos ou navigation vers d'autres vidéos hébergées sur le même site) ne relève ni du contrôle ni de la responsabilité de SAP.

Fonctionnalités Beta et expérimentales

Les fonctionnalités expérimentales ne font pas partie des éléments officiellement fournis par SAP et garantis pour les versions à venir. Cela signifie que les fonctionnalités expérimentales peuvent être modifiées par SAP à tout moment pour quelle que raison que ce soit, sans préavis. Les fonctionnalités expérimentales ne sont pas conçues pour être utilisées en production. Vous ne pouvez pas faire la démonstration, tester, examiner, évaluer ou utiliser d'une quelconque autre manière les fonctionnalités expérimentales dans un environnement productif ou avec des données n'ayant pas été suffisamment sauvegardées.
Le but des fonctionnalités expérimentales est d'obtenir rapidement des avis afin que les clients et partenaires puissent influencer le produit futur. En partageant votre avis (par exemple sur SAP Community), vous acceptez que les droits de propriété intellectuelle des contributions ou œuvres dérivées constituent la propriété exclusive de SAP.

Exemple de code

Les codes et/ou fragments de code ne sont que des exemples. Ils ne sont pas destinés à une utilisation en production. L'exemple de code est utilisé uniquement pour mieux expliquer et visualiser les règles de syntaxe. SAP ne garantit pas l'exactitude ni l'exhaustivité de l'exemple de code. SAP ne saurait être tenue responsable des erreurs ou dommages causés par l'utilisation dudit exemple de code sauf si de tels dommages étaient causés par une négligence grave ou une faute intentionnelle de SAP.

Langage sans préjugés

SAP soutient une culture de diversité et d'inclusion. Chaque fois que cela est possible, nous utilisons un langage impartial dans notre documentation pour faire référence aux personnes de toutes cultures, ethnies, genres et capacités.

www.sap.com/contactsap

© 2023 SAP SE ou société affiliée SAP. Tous droits réservés.

Toute reproduction ou communication de la présente publication, même partielle, par quelque procédé et à quelque fin que ce soit, est interdite sans l'autorisation expresse et préalable de SAP SE ou d'une société affiliée SAP. Les informations du présent document sont susceptibles d'être modifiées sans préavis.

Certains logiciels commercialisés par SAP SE et ses distributeurs contiennent des composants logiciels qui sont la propriété d'éditeurs tiers. Les spécifications des produits peuvent varier d'un pays à l'autre.

Les informations du présent document sont fournies par SAP SE ou par une société affiliée SAP uniquement à titre informatif, sans engagement ni garantie d'aucune sorte. SAP SE ou ses sociétés affiliées ne pourront en aucun cas être tenues responsables des erreurs ou omissions relatives à ces informations. Les seules garanties fournies pour les produits et les services de SAP SE ou d'une société affiliée SAP sont celles énoncées expressément à titre de garantie accompagnant, le cas échéant, lesdits produits et services. Aucune des informations contenues dans le présent document ne saurait constituer une garantie supplémentaire.

SAP et tous les autres produits et services SAP mentionnés dans ce document, ainsi que leurs logos respectifs, sont des marques commerciales ou des marques déposées de SAP SE (ou d'une société affiliée SAP) en Allemagne ainsi que dans d'autres pays. Tous les autres noms de produit et service mentionnés sont des marques commerciales de leurs sociétés respectives.

Veuillez consulter https://www.sap.com/france/about/legal/trademark.html pour plus d'informations sur les marques déposées.

THE BEST RUN SAP