Getting Started With Association Rules
Automated Analytics User Guide
# Content

1. **About this Document** .......................................................... 3

2. **Association Rules Definitions** ................................................ 5

3. **Introduction to Sample Files** ................................................ 10

4. **Modeling with Association Rules** ............................................. 11
   4.1 Creating Association Rules .................................................. 11
   4.2 Step 1 - Selecting the Data .................................................. 11
      Selecting a Data Source ...................................................... 11
      Describing the Data .......................................................... 15
      Selecting Events Data ....................................................... 16
      Describing Events Data ...................................................... 16
   4.3 Scenario 1: Standard Modeling with Association Rules .............. 17
      Step 2 - Defining the Modeling Parameters ................................ 17
      Step 3 - Generating the Model ............................................. 22
      Step 4 - Analyzing and Understanding the Generated Model ......... 23
   4.4 Scenario 2: Modeling Using Sequence Mode ............................. 31
      Step 2 - Defining the Modeling Parameters ................................ 32
      Step 3 - Generating the Model ............................................. 34
      Step 4: Model Debriefing .................................................... 34
      Step 5: Using the Model ..................................................... 36

5. **Modeling Using KxShell Script** ............................................. 40
   5.1 Learning the Model .......................................................... 44
   5.2 Applying the Model .......................................................... 44
   5.3 Saving the Model ............................................................. 45
   5.4 Opening an Existing Model .................................................. 45
   5.5 Summarizing the KxShell script ........................................... 45
1 About this Document

Who Should Read this Document

This document is addressed to people who want to evaluate or use the Automated Analytics Modeler and in particular the Association Rules feature.

Prerequisites

Before reading this guide, you should read chapters 2 and 3 of the Classification, Regression, Segmentation and clustering Scenarios – Automated Analytics User Guide that present respectively:

- An introduction to the Automated Analytics.
- The essential concepts related to use of the Automated Analytics features.

No prior knowledge of SQL is required to use data manipulation -only knowledge about how to work with tables and columns accessed through ODBC sources. Furthermore, users must have “read” access on these ODBC sources.

To use the Java graphical interface, users need write access on the tables KxAdmin and ConnectorsTable, which are used to store representations of data manipulations.

What this Document Covers

This document introduces you to the main functionalities of the Association Rules feature.

Association Rules generates association rules. Association rules provide clear and useful results, especially for market basket analysis. They bring to light the relations between products or services and immediately suggest appropriate actions. Association rules are used in exploring categorical data, also called items.

The strengths of Association Rules are:

- to produce clear and understandable results,
- to support unsupervised data mining (no target attribute),
- to explore very large data sets thanks to its ability to first generate rules on parts of the data set before aggregating them (exploration by chunks),
- to generate only the more relevant rules (also called primary rules).
How to Use this Document

Organization of this Document

This document is subdivided into five chapters.

The current chapter, About this Document, serves as an introduction to the rest of the guide. This is where you will find information pertaining to the reading of this guide, and information that will allow you to contact us.

Chapter 2, Association Rules Definitions, provides definitions relative to the Association Rules that will allow you a better understanding of the feature functionalities.

Chapter 3, Introduction to Sample Files, presents the data files provided to start with Association Rules.

Chapter 4, Modeling with Association Rules, presents the Association Rules feature. This chapter is organized in five parts:

- The first part presents the first step of the modeling process, that is the data selection and description.
- The second part presents the standard use of the Association Rules feature.
- The third part describes how to use sequences when modeling with Association Rules.
- The fourth part describes how to save or apply an Association Rules model.
- The last part provides you with a KxShell script based on the sample files described in chapter 3.

Chapter 5, Modeling Using KxShell Script, details a KxShell script that can be used with the provided sample files.

A detailed table of contents located at the beginning of the guide and cross-references throughout the document allow you to find the information that you need, quickly and easily.

If you want more information on the essential concepts of modeling data, read the Classification, Regression, Segmentation and Clustering Scenarios – Automated Analytics User Guide.

Full Documentation

2 Association Rules Definitions

This part gives some definitions relative to the Association Rules.

Session

A session is a set of transactions identified by a unique key; for example, all the purchases done at one time by a single customer.

Transaction

A transaction is defined by:

- a unique key,
- the key of the related session,
- an attribute, called an item.

Itemset

A group (or a set) of items is called an itemset.

Association Rule

An Association Rule is an implication relation of the form $X \Rightarrow Y$. The rule means: if the attribute $X$ is present in a session, then the attribute $Y$ is present too. Two measures allow qualifying the quality of the rule: the Support and the Confidence.

Antecedent

$X$ is called the antecedent of the rule. The antecedent can be composed of an item or an itemset, for example $X$ can be the set \{A,B,C\}.
**Consequent**

Y is called the consequent of the rule. The consequent is composed of only one item, for example Y can be the item \( D \).

**Support**

The Support of a rule is a measure that indicates the number of sessions that verify the rule. For instance the number of sessions, that contains the itemset \{A,B,C\} and the item D.

**Confidence**

The Confidence of a rule is a measure that indicates the percentage of sessions verifying the consequent among those verifying the antecedent. For instance the number of sessions containing the item D, among the ones containing the itemset \{A,B,C\} (see graph below).

![Venn Diagram](image)

The formula used to calculate the confidence is:

\[
\text{Confidence} = \frac{\text{Support(ABCD)}}{\text{Support(ABC)}}
\]
Lift

The Lift of a rule is a measure that indicates the chances of finding the consequent by using the antecedent compared with the chances of randomly finding the consequent. A value greater than 1 indicates that using the antecedent increases your chances to find the consequent.

Predictive Power (KI)

The predictive power (KI) indicator is the quality indicator of the models generated using Automated Analytics. This indicator corresponds to the proportion of information contained in the consequent that the antecedent is able to explain.

The following diagram graphically represents:

- an association rule (black curve), that is the chances of the antecedent rule to allow finding the expected consequent,

compared with:

- a random selection (red curve), that is the chances of a randomly selected item in the data set to be the expected consequent,
- and a perfect selection (green curve), that is the case when all the first selected items are the expected consequent.

In this diagram, the KI indicator corresponds to “the area found between the curve of the rule generated and that of the random selection” divided by “the area found between the curve of the perfect selection and that of the random selection”. As the curve of the generated rule approaches the curve of the perfect selection, the value of KI approaches 1.

Predictive Power in Relation to the Classic Indicators

This section defines the mathematical relation of the predictive power to other standard indicators in the Association Rules Mining area: Rule Support, Confidence and Lift.
Assuming we have an association rule $R$ with the shape: $R: A \Rightarrow B$ (EQ 1) where $A$ is the antecedent and $B$ the consequent. The following figure, named Venn diagram, represents the association rule $R$ defined above.

Associated with the association rule $R$, we make the following assumptions about the variety of the supports to simplify the following computations (all support measures are represented as proportions).

$$\text{Supp}(A) = a + c$$  
$$\text{Supp}(B) = b + c$$  
$$\text{Supp}(R) = c$$  
$$\text{Supp}(\overline{A}) = 1 - (a + c)$$  
$$\text{Supp}(\overline{B}) = 1 - (b + c)$$  
$$\text{Supp}(A \cap B) = d = 1 - (a + b + c)$$  
$$a + b + c + d = 1 \quad (EQ \ 2)$$

where

$$0 \leq a \leq 1$$  
$$0 \leq b \leq 1$$  
$$0 \leq c \leq 1$$  
$$0 \leq d \leq 1$$
The Venn diagram presented above could be mapped to the EQ 2 in a confusion matrix:

<table>
<thead>
<tr>
<th>Actual Consequent</th>
<th>Predicted (from Antecedent)</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>c</td>
<td>b</td>
<td>Supp(B)</td>
</tr>
<tr>
<td>0</td>
<td>a</td>
<td>d</td>
<td>Supp(Not B)</td>
</tr>
<tr>
<td>Supp(A)</td>
<td>Supp(Not A)</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Note

The different supports represent the support of a set with respect to the whole population.

Using EQ 2 and the Confusion Matrix representation of the KI, we can extract the following definition.

\[
KI = \frac{c - (a + c)(b + c)}{(b + c)\left(1 - (b + c)\right)}
\]

\[
KI = \frac{c - (a + c)(b + c)}{(b + c)(a - (b + c))}
\]

\text{Note}

Using EQ 3, predictive power (KI) could be expressed in terms of:

**Rule Support:**

\[
KI = \frac{\text{Supp}(R) - \text{Supp}(A)\ \text{Supp}(B)}{\text{Supp}(B)\left(1 - \text{Supp}(B)\right)}
\]

\text{(EQ 4)}

**Confidence:**

\[
KI = \frac{\text{Supp}(A) \left[\text{Conf}(R) - \text{Supp}(B)\right]}{\text{Supp}(B)\left(1 - \text{Supp}(B)\right)}
\]

\text{(EQ 5)}

**Lift:**

\[
KI = \frac{\text{Supp}(A) \left(\text{Lift}(R) - 1\right)}{1 - \text{Supp}(B)}
\]

\text{(EQ 6)}
3 Introduction to Sample Files

You can test the Association Rules feature using the sample files available in the folder `Samples/KAR` located:

- for Windows, in the folder `Program Files\SAP BusinessObjects Predictive Analytics <version number>\Automated`
- for UNIX, in the folder where you have decompressed the archive file (that is .tar.Z or .tar.gz).

The data set contains a single day of Web traffic from an E-commerce site in December 1999. The site content was served by a Broadvision server, but no "cookies" or login was required, making the sessions effectively anonymous.

<table>
<thead>
<tr>
<th>File</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>website_references.csv</code></td>
<td>list of sessions and binary purchase target (50581 rows)</td>
</tr>
<tr>
<td><code>website_references_desc.csv</code></td>
<td>description for <code>website_references.csv</code></td>
</tr>
<tr>
<td><code>website_transactions.csv</code></td>
<td>log of files requested from Broadvision server (532860 rows)</td>
</tr>
<tr>
<td><code>website_transactions_desc.csv</code></td>
<td>description for <code>website_transactions.csv</code></td>
</tr>
</tbody>
</table>

The parameters of the model used in this guide are based on these sample files.

**Note**
A Read-Me file is provided with the sample files describing the data sets and how to use them to test Association Rules feature.

### Additional Sample Files

Another set of sample files is available in the folder `Samples\KAR` with a readme text file describing how to use the following samples:

- `customers_references.txt`
- `customers_references_desc.txt`
- `customers_transactions.txt`
- `customers_transactions_desc.txt`
4 Modeling with Association Rules

4.1 Creating Association Rules

Association Rules generates association rules. Association rules provide clear and useful results, especially for market basket analysis. They bring to light the relations between products or services and immediately suggest appropriate actions. Association rules are used in exploring categorical data, also called items.

The strengths of Association Rules are:

- to produce clear and understandable results,
- to support unsupervised data mining (no target attribute),
- to explore very large data sets thanks to its ability to first generate rules on parts of the data set before aggregating them (exploration by chunks),
- to generate only the more relevant rules (also called primary rules).

1. Select Start Programs SAP Business Intelligence SAP BusinessObjects Predictive Analytics Desktop SAP BusinessObjects Predictive Analytics opens up.
2. Select the Association Rules feature in the Modeler section.

4.2 Step 1 - Selecting the Data

4.2.1 Selecting a Data Source

After selecting the type of model that you want to generate, you must select the data source that you want to use as the training data set.

An Association Rules Reference data set must have a single variable unique primary key. If the primary key is non-unique or spread out over several variables, Association Rules will not function properly.

1. On the screen Select a Data Source, select the Data Type to be used (Text file, ODBC, ...).
   For our example, select the option Text File.
2. Click the Browse button.
   The Data Source Selection dialog opens.
3. Select the file you want to use, then click OK.
   The name of the file will appear in the Data Set field.
   For our example, select the file website_references.csv.
4. Click Next.

4.2.1.1 Data Sources Supported

Automated Analytics supports the following data sources:

- **Text files** (also called flat files) in which the data are separated by a delimiter, such as commas in `.csv` (Comma Separated Value) files.

  **Restriction**
  
  When accessing data in `.csv` files, Automated Analytics only supports CR + LF (common on Microsoft Windows) or LF (common on UNIX) for line breaks.

- **Database management systems** that can be accessed using ODBC.

  **Note**
  
  For the list of supported ODBC-compatible sources, see the SAP Product Availability Matrix (PAM) at http://service.sap.com/sap/support/pam.

  For more information about using SAP HANA, see the related information below.

  To configure Automated Analytics modeling tools to access data in your database management system, refer to the guide Connecting your Database Management System on Windows or Connecting your Database Management System on UNIX.

- **SAS files**

4.2.1.2 SAP HANA as a Data Source

You can use SAP HANA databases as data sources for all types of modeling analyses in Modeler: Classification/Regression, Clustering, Time Series, Association Rules, Social, and Recommendation.

| SAP HANA tables or SQL views | found in the Catalog node of the SAP HANA database |
### All types of SAP HANA views

An SAP HANA view is a predefined virtual grouping of table columns that enables data access for a particular business requirement. Views are specific to the type of tables that are included, and to the type of calculations that are applied to columns. For example, an analytic view is built on a fact table and associated attribute views. A calculation view executes a function on columns when the view is accessed.

**Restriction**

- Analytic and calculation views that use the variable mapping feature (available starting with SAP HANA SPS 09) are not supported.
- You cannot edit data in SAP HANA views using Automated Analytics.
- SAP HANA views are not supported in Data Manager.

### Smart Data Access virtual tables

Thanks to Smart Data Access, you can expose data from remote sources tables as virtual tables and combine them with HANA regular tables. This allows you to access data sources that are not natively supported by the application, or to combine data from multiple heterogeneous sources.

**Caution**

To use virtual tables as input data sets for training or applying a model or as output data sets for applying a model, you need to check that the following conditions are met:

- The in-database application mode is not used.
- The destination table for storing the predicted values exists in the remote source before applying the model.
- The structure of the remote table, that is the column names and types, must match exactly what is expected with respect to the generation options; if this is not the case an error will occur.

**Caution**

In Data Manager, use virtual tables with caution as the generated queries can be complex. Smart Data Access may not be able to delegate much of the processing to the underlying source depending on the source capabilities. This can impact performance.
Prerequisites

You must know the ODBC source name and the connection information for your SAP HANA database. For more information, contact your SAP HANA administrator.

In addition to having the authorizations required for querying the SAP HANA view, you need to be granted the SELECT privilege on the _SYS_BI schema, which contains metadata on views. Please refer to SAP HANA guides for detailed information on security aspects.

4.2.1.3 To Select an SAP HANA Table or SQL View as Data Source

1. On the Select a Data Source screen, select Data Base in the Data Type list.
2. Click the Browse button left of the Folder field.
3. In the Data Selection dialog, select the SAP HANA database you want to use in the Select Source Folder for Data list.
4. If this is the first connection to this database, enter the login information in the User and Password fields and click Connect.
   
   The content of the database is displayed in the left part of the window.
5. Open the Catalog node, select the schema containing the table or SQL view you want to use as the data source for your model and click OK.
6. Click the Browse button left of the Data Set field.
7. In the Data Selection dialog, select the table or SQL view you want to use.
   You can filter the table list by entering part of the table or SQL view name in the field located below the list.
8. Click OK.

4.2.1.4 To Select an SAP HANA View

You can select an SAP HANA analytic or calculation view as a data source for your model.

1. On the Select a Data Source screen, select Data Base in the Data Type list.
2. Click the Browse button left of the Folder field and select the SAP HANA database you want to use in the Select Source Folder for Data list.
3. If this is the first connection to this database, enter the login information in the User and Password fields and click Connect.
   
   The content of the database is displayed in the left part of the window.
4. Open the Content node. Use the tree to browse the packages and select the one containing the SAP HANA view you want to use as the data source for your model and click OK.
5. Click the Browse button left of the Data Set field.
6. In the Data Selection dialog, select the view you want to use.
   You can filter the view list by entering part of the view name in the field located below the list.
7. Click **OK**.

If the SAP HANA view you have selected requires you to specify some values, the **User Values** window opens when you click **View Data** or **Next** on the **Select a Data Source** screen. The variables for which you need to enter a value are listed on the left side of the panel. The following table details how the different variables are identified in the list. In this topic, the term variable is used for both variables and input parameters.

<table>
<thead>
<tr>
<th>Type of variable</th>
<th>Noted as</th>
<th>Signaled by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandatory variable</td>
<td>mandatory</td>
<td>A red star</td>
</tr>
<tr>
<td>Optional variable</td>
<td>optional</td>
<td></td>
</tr>
<tr>
<td>Variable with a default value</td>
<td>default value</td>
<td>A green check mark</td>
</tr>
<tr>
<td>Variable with a value set by the user</td>
<td></td>
<td>A green check mark</td>
</tr>
</tbody>
</table>

Check the box **Show mandatory only** to display only mandatory variables in the list.

8. In the list, select the variable you want to set a value for.

The field allowing you to set the value for the variable is displayed on the right side of the panel. Depending on the type of value required, you may see different types of fields. For example, if a single value is required, you see a list of operators to choose from and a field to enter a value. If a range of values is required, you see two fields for the range bounds and a list of operators to choose from.

9. Some variables have a list of preset possible values. In that case, the list of values for the current variable is displayed on the right side of the panel. Select the value you want to set for this variable; it will be displayed in the value field.

10. Use the **Restore** button to reset the value of the current variable to its default value.

11. Once you have entered all required values, click **OK**.

Once you have accessed an SAP HANA view, the values you have set will be remembered and used as default when you access this view again during the modeling process.

### 4.2.2 Describing the Data

In order for the application features to interpret and analyze your data, the data must be described. To put it another way, the description file must specify the nature of each variable, determining their:

- **Storage format**: number (**number**), character string (**string**), date and time (**datetime**) or date (**date**).
- **Type**: **continuous**, **nominal**, **ordinal** or **textual**.

For more information about data description, see the Types of Variables and Storage Formats in Classification, Regression, Segmentation and Clustering Scenarios – Automated Analytics User Guide.

To describe your data, you can:

- Either use an existing description file, that is, taken from your information system or saved from a previous use of the application features,
- Or create a description file using the Analyze option, available to you in the application. In this case, it is important that you validate the description file obtained. You can save this file for later re-use.
Caution

The description file obtained using the Analyze option results from the analysis of the first 50 lines of the initial data file. In order to avoid all bias, we encourage you to mix up your data set before performing this analysis.

Using a description file:

1. On the screen Data Description, use the Browse button to select the description file. For our example, select the file website_references_desc.csv.
2. Click the button Open Description.
3. Select the description file and click OK.
4. Click the Next button.

Analyzing the data set:

5. On the screen Data Description, click the Analyze button.
6. To verify the description you obtained, click the button View Data and take a look at the data set beginning.
7. Click the Next button.

4.2.3 Selecting Events Data

The screen Events Data Source lets you specify the data source to be used as the Transaction data set.

1. Select the Data Type of your data source (Text Files, ODBC, ...).
   For our example, select the option Text Files.
2. In the Folder field, specify the folder where your data source is stored.
3. In the Events field, specify the name of your data source.
   In our example, select the file website_transactions.csv.
4. Click the Next button.

4.2.4 Describing Events Data

The screen Events Data Description lets you describe your Transaction data, offering you the same options as the screen Data Description. For Association Rules to function properly, there must be a variable in the Transaction data set that is the same as the primary key declared for the Reference data set, referred to as a “Join Column”. The name of the variable can be different, but the storage and value must be the same. The values of this variable need not be unique, since each Reference key can have 0, 1, or several associated transactions.

In addition to a suitable join column, the Transaction data set must have at least one item variable. The item variable will be used by Association Rules to build the association rules.

For detailed procedures on how to set parameters on this screen, see Describing the Data.

There are two ways to describe the events data, either by using a description file, or by letting the application analyze the first hundred lines of the data set.
Using a description file:
1. On the screen *Events Data Description*, select the description file.
   
   For our example, select the file `website_transactions_desc.csv`.
2. Click the button *Open Description*.
   
   The data description is displayed.
3. Select the description file and click *OK*.
4. Click *Next*.

Analyzing the data set:
5. On the screen *Events Data Description*, click the *Analyze* button.
6. To verify the description you obtained, click the button *View Data* and take a look at the data set beginning.
7. Click *Next*.

### 4.3 Scenario 1: Standard Modeling with Association Rules

#### 4.3.1 Step 2 - Defining the Modeling Parameters

##### 4.3.1.1 Setting ASSOCIATION RULES Parameters

The screen Association Rules Extraction Parameters enables you to set Association Rules parameters by:

- Joining your reference data with your transaction data
- Setting the Specific Modeling Parameters

1. On the screen *Association Rules Extraction Parameters*, select the join column for both the log and reference data sets.
   
   For our example, select the column SessionID in both data sets.
2. Select the *Item Column*.
   
   For our example, select the column Page.
3. Set the *Minimum Support*.
   
   For our example, enter 1.
4. Click *Next*.

##### 4.3.1.2 Understanding Association Rules Parameters

The Association Rules parameters allow you to set the format of the rules you want to get by setting the following indicators:

- the minimum support
• the minimum confidence
• the maximum length

Minimum Support

The support of a rule is the number of records verifying the rule. With a rule of the form $X \Rightarrow Y$, the support is the count of records containing the itemset $X$ and the attribute $Y$.

The default value is 10. If the value is superior to 10, it represents the number of sessions. If the value is between 0 and 10, it represents a percentage of the number of sessions (default = 10, required value > 0).

Minimum Confidence

The confidence of a rule is the percentage of records verifying the consequent of the rule among those verifying the antecedent of the rule. With a rule of the form $X \Rightarrow Y$, the confidence is the count of records containing the itemset $X$ and the attribute $Y$ in relation to the count of records containing only the itemset $X$.

The default value is 0.5. This parameter requires a value between 0 and 1.

Maximum Length

The maximum length of a rule is its total number of items, including the antecedent and the consequent. It is called the cardinality of the rule.

The default value is 4. This parameter requires a value higher or equal to 2.

4.3.1.3 Understanding Advanced Parameters

You can set the association rules advanced parameters by clicking the Advanced button located on the right bottom corner of the panel Association Rules Extraction Parameters.

Chunk Size

The option Chunk Size allows you to set a minimal number of sessions for which Association Rules will generate temporary rules.

In a standard use, that is when the chunk size is 0, Association Rules imports all the transactions in memory before generating the association rules corresponding to the model parameters (minimum support, minimum confidence, maximum rule length). With large data sets, this method can be very expensive in time and memory.
To gain speed and memory space, Association Rules offers the option to import and generate rules from pieces of the events file. In details, Association Rules will load in memory a number of sessions equivalent to the value of the “Chunk Size” parameter. Temporary Association Rules are generated on this set of sessions. Then, memory is cleaned from these sessions and Association Rules will load the next sessions with a number equivalent to the value of the “Chunk Size” parameter. This operation will be repeated until the end of the Events data file. At the end of the learning step, these rules are joined and their statistics are updated for the entire data set.

The more sessions you have to analyze, the more interesting the option Chunk Size becomes; although the number of rules and of frequent found itemsets can differ depending on the method you select. However, these differences are not significant if the chunks size is large enough. The table below describes the utility of this option depending on the data sets specifications.

<table>
<thead>
<tr>
<th>Chunk Size Use Recommendation</th>
<th>Low Number of Sessions</th>
<th>High Number of Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Number of Transactions in the Session</td>
<td>Not Needed</td>
<td>Strongly Recommended</td>
</tr>
<tr>
<td>High Number of Transactions in the Session</td>
<td>Not Needed</td>
<td>Strongly Recommended</td>
</tr>
</tbody>
</table>

As an indication, the chunk size should be roughly 10% of the total number of sessions in your data set (that is 10 to 15 chunks depending on the data set size). You can start with 10% and modify this number to improve your results. If your transactions file contains less than 5000 rows, you should use a no-chunk strategy.

**Caution**

The transactions in the events data set must be grouped by sessions, meaning that the Events file must be sorted by sessions.

However, the task of splitting the sessions in chunks can be delegated to the application through the option Guess Chunk Size which will automatically compute the chunks’ size in order to obtain ten of them.

**Skip Derived Rules / Skip Reducible Itemsets**

The option Skip Derived Rules indicates to Association Rules to generate only the primary rules and thus skip the derived rules. For instance, given R1 = X => Z and R2 X,Y =>Z two association rules, R1 is called a primary rule and R2 a derived rule. The default value is True.

**Filters on Consequent Items**

Before starting the learning process, you can select which items you want to see as consequents in the rules generated by Association Rules by defining either the consequents to keep or the ones to exclude from the modeling.
4.3.1.3.1 Selecting the Items to Keep

1. Select the option *Items to Include in the Consequent*.
   This option is selected by default.

2. In the text field located at the bottom of the panel (1) enter the value of the item you want to see appear in the rules.
3. Click the + button located on the right of the field. The item appears in the list above (2).
4. Repeat steps 2 and 3 for all the items you want to include.
5. Click the *Validate* button to save the list and go back to the panel Association Rules Parameters *Settings*.

4.3.1.3.2 Excluding Items from the Consequents

1. Select the option *Items to Exclude in the Consequent*.
2. In the text field located at the bottom of the panel enter the item you do not want to see appear as consequent in the rules.
3. Click the + button located on the right of the field. The item appears in the list above.
4. Repeat steps 2 and 3 for all the items you want to exclude.
5. Click the *OK* button to save the list and go back to the panel Association Rules Parameters *Settings*.

4.3.1.3.3 Removing Items

1. Select the items you want to remove (use the Ctrl key to select several items).
2. Right-click the list. A contextual menu is displayed.
3. Select the option *Remove this (these) item(s).*
4.3.1.3.4 Renaming an Item

1. Select the item you want to modify.
2. Right-click the list.
   A contextual menu is displayed.
3. Select the option Rename this item.
   The item value is displayed in the text field.
4. Modify the value as needed.
5. Click the + button to validate the change.
   To sort a list alphabetically, check the box Alphabetic Sort if you want to see the consequent listed in an alphabetical order.

4.3.1.3.5 Checking Modeling Parameters

The Summary of Modeling Parameters screen allows you to check the modeling parameters just before generating the model.

- The name of the model is filled automatically. It corresponds to the name of the item variable, followed by the underscore sign (“_”) and the name of the Reference file without its file extension.
- The Autosave button allows you to activate the feature that will automatically save the model once it has been generated. When the autosave option is activated, a green check mark is displayed on the Autosave button.

4.3.1.3.6 Activating the Autosave Option

The Model Autosave panel allows you to activate the option that will automatically save the model at the end of the generation process and to set the parameters needed when saving the model.

To activate the option, proceed as follows:

1. In the Summary of Modeling Parameters panel, click the Autosave button.
   The Model Autosave panel is displayed.
2. Check Enable Model Autosave.
3. Set the parameters listed in the following table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Name</td>
<td>This field allows you to associate a name with the model. This name will then appear in the list of models to be offered when you open an existing model.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Description</td>
<td>This field allows you to enter the information you want, such as the name of the training data set used, the polynomial degree or the performance indicators obtained. This information could be useful to you later for identifying your model. Note that this description will be used instead of the one entered in the Summary of Modeling Parameters panel.</td>
</tr>
<tr>
<td>Data Type</td>
<td>This list allows you to select the type of storage in which you want to save your model. The following options are available: - Text Files, to save the model in a text file. - Data Base, to save the model in a database. - Flat Memory, to save the model in the active memory.</td>
</tr>
<tr>
<td>Folder</td>
<td>Depending upon which option you selected, this field allows you to specify the ODBC source, the memory store or the folder in which you want to save the model.</td>
</tr>
<tr>
<td>File/Table</td>
<td>This field allows you to enter the name of the file or table that is to contain the model. When saving the model as a text file, you must enter one of the following format extensions .\text{txt} (text file in which the data is separated by tabs) or .\text{csv} (text file in which the data is separated by commas).</td>
</tr>
</tbody>
</table>

4. Click **OK**.

### 4.3.2 Step 3 - Generating the Model

Once the modeling parameters are defined, you can generate the model.

1. **On the Summary of Modeling Parameters screen, click the Generate button.**

   The **Training the Model** screen opens. The model is being generated. A progress bar allows you to follow the process.

2. **If the Autosave option has been activated in the Summary of Modeling Parameters panel, a warning message is displayed at the end of the learning process confirming that the model has been saved.**

3. **Click Close.**

### 4.3.2.1 Following the Generating Process

There is two ways for you to follow the progress of the generation process:

- The Progress Bar displays the progression for each step of the process. It is the screen displayed by default.
- The Detailed Log displays the details of each step of the process.

To display the detailed log, click the **Show Detailed Log** button.

To display the progression bar, click the **Show Progression** button. The progression bar screen appears.

To stop the learning process:
1. Click the Stop Learning Process button.
2. Click Previous. The Summary of Modeling Parameters screen appears.
3. Go back to the Checking Modeling Parameters section.

Visualizing the Model Results

At the end of the generation process, a summary of the model results appears.

For more information on the model summary, go to step 4, Analyzing and Understanding the Model (see “Step 4 - Analyzing and Understanding the Generated Model”).

4.3.3 Step 4 - Analyzing and Understanding the Generated Model

Automated Analytics allows you to analyze and understand the model generated thanks to a suite of plotting tools describing:

- all the items treated by the model,
- the rules generated by the model.

Model Debriefing

To display the Model Debriefing, on the Using the Model menu, select the Model Overview option. The Model Overview screen opens.

**Note**

If you have built more than one model in the same session, the debriefing for all the models will be displayed on this screen, sorted by Date of Build.

Understanding the Model Debriefing

The Model Overview screen is composed of four sections detailing information on:

- the model,
- the transformation,
- the parameters,
- the modeling results.
The Model

This section details the following information:

<table>
<thead>
<tr>
<th>Name</th>
<th>Significance</th>
<th>For example...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Name of the model. It is generated by using the item variable name and the</td>
<td>Page_website_references</td>
</tr>
<tr>
<td></td>
<td>reference data set name.</td>
<td></td>
</tr>
<tr>
<td><strong>Data Set</strong></td>
<td>Name of the Reference data set used for the model.</td>
<td>website_references.csv</td>
</tr>
<tr>
<td><strong>Initial Number of Variables</strong></td>
<td>Total number of variables in the Reference data set.</td>
<td>1</td>
</tr>
<tr>
<td><strong>Number of Selected Variables</strong></td>
<td>Number of variables used to build the model.</td>
<td></td>
</tr>
<tr>
<td><strong>Building Date</strong></td>
<td>Date and time when the model was built.</td>
<td>2012-04-20 15:45:23</td>
</tr>
<tr>
<td><strong>Learning Time</strong></td>
<td>Time needed to build the model.</td>
<td>1s</td>
</tr>
<tr>
<td><strong>Engine Name</strong></td>
<td>Name of the feature used to build the model.</td>
<td>Association Rules</td>
</tr>
</tbody>
</table>

The Training Parameters

This section summarizes the parameters used to build the model. The following table details the provided information:

<table>
<thead>
<tr>
<th>Name</th>
<th>Significance</th>
<th>For example...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimum Support</strong></td>
<td>The minimum number of sessions to validate a rule.</td>
<td>10%</td>
</tr>
<tr>
<td><strong>Minimum Confidence</strong></td>
<td>The minimum percentage of sessions verifying the consequent among those verifying the antecedent.</td>
<td>50%</td>
</tr>
<tr>
<td><strong>Maximum Length</strong></td>
<td>The maximum length of the rule.</td>
<td>4</td>
</tr>
</tbody>
</table>

The Data Sets Parameters

This section summarizes the parameters of the data sets used to build the model. The following table details the provided information:

<table>
<thead>
<tr>
<th>Name</th>
<th>Significance</th>
<th>For example...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transaction ID Column</strong></td>
<td>name of the join key for the events data set.</td>
<td>SessionID</td>
</tr>
<tr>
<td><strong>Reference ID Column</strong></td>
<td>name of the join key for the reference data set.</td>
<td>SessionID</td>
</tr>
<tr>
<td><strong>Item Column</strong></td>
<td>name of the column containing the items.</td>
<td>Page</td>
</tr>
</tbody>
</table>

The Results

This section displays information on the rules generated.
The following table details the provided information:

<table>
<thead>
<tr>
<th>Name</th>
<th>Significance</th>
<th>For example...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Sessions Processed</strong></td>
<td>The number of sessions treated by the model.</td>
<td>245</td>
</tr>
<tr>
<td><strong>Number of Rules Found</strong></td>
<td>The number of rules found by the model.</td>
<td>61</td>
</tr>
<tr>
<td><strong>Number of Items Found</strong></td>
<td>The number of items found by the model.</td>
<td>481</td>
</tr>
<tr>
<td><strong>Number of Frequent Item Sets</strong></td>
<td>The number of item sets whose support is superior to the minimum support.</td>
<td>731</td>
</tr>
<tr>
<td><strong>Transactions Found</strong></td>
<td>The number of transactions found by the model</td>
<td>2500</td>
</tr>
</tbody>
</table>

⚠️ Caution

The Frequent Item sets information should not be taken into account if the Chunk learning strategy has been used. In this particular case, the information may be underestimated.

**Association Rules Exploration**

On the Using the Model menu, select the Association Rules Exploration option. The Association Rules Exploration screen opens. This screen allows you to search through all the rules found by the association rules engine thanks to filtering options and allows you to display the rules as a graph.

**Setting the Search Parameters**

There are three types of filters available:

- by Antecedent, where you select the antecedents you want to see in the rules
- by Consequent, where you select the consequents you want to see in the rules
- by numerical filters:
  - Rule Support (in number of sessions or in percentage)
  - Antecedent Support (in number of sessions or in percentage)
  - Consequent Support (in number of sessions or in percentage)
  - KI (predictive power)
  - Confidence
  - Lift

You can either apply only one filter or combine several to refine your search.
To Set a Filter by Antecedent and/or Consequent, set the filter(s) you want to apply:

- by antecedent: select one or more antecedents in the list,
- by consequent: select one or more consequents in the list.

If you filter both by antecedent and consequent be careful not to select the same item in both lists.

### 4.3.3.1 Setting a Numerical Filter

1. Select an indicator in the list (Confidence, Rule Support, Size, Lift, KI, Antecedent Support). (1)
2. Select a comparison operator. (2)
3. Enter the value of comparison in the text field. (3)
4. Click the > button to add the filter to the filter list (4).

Example: KI >= 0.8

#### Note

1. To start the search, once the filters are set, click the Search button. The tab rule(s) found displays the number of rules found and the list of these rules. You can sort them by any column: just click the select column header.

#### Caution

If you click the Search button without selecting a filter, all the rules will be displayed. When the number of rules is very high, displaying all of them can take a very long time.

2. To reset the search engine, click the Reset button.

### 4.3.3.2 Displaying the Search Results

The search results are displayed in the tab **rules found**.

To Select the Rules to Display

1. Check or uncheck the Activated box next to the rules you want to display/hide.
2. Right-click one of the table fields. A contextual menu is displayed.

3. Choose one of the offered options.

To Select the Columns to Display

4. Right-click the table header. A contextual menu is displayed.

5. Check the columns you want to see displayed in the result list. The selected columns automatically appear in the list.

Note

To sort the results by column, click the selected column header.
4.3.3.3 Displaying the Rules as a Graph

The tab **Graph View** allows you to display the rules as a graph.

To display the Association Rules as a graph:

- Either filter the rules found by **Association Rules** and then display the corresponding graph: On the tab .. rule(s) found, click the button **Display Filtered Rules as a Graph**.
- Or select the antecedent(s) and/or consequent(s) you want to see in the graph (see section Understanding the Graph below).

To Add an Item to the Graph:
1. Drag and drop the selected item from the list of items to the graph section.
   A contextual menu is displayed.
2. Select if you want to display the rules where the item is part of the rule antecedent, the rule consequent or both:
   - if you select **Antecedent**, only the rules where the antecedent contains the selected item will be displayed,
   - if you select **Consequent**, only the rules where the consequent contains the selected item will be displayed,
   - if you select **Antecedent or Consequent**, only the rules where the antecedent or the consequent contains the selected item will be displayed.

4.3.3.4 Understanding the Graph Panel

The tab **Graph View** is split in two parts:

- (1) on the left, the list of the items used in at least one rule, either as an antecedent or as a consequent.
- (2) on the right, the graph.
The following table sums up the visual elements used in the graph:

<table>
<thead>
<tr>
<th>The element...</th>
<th>...represents</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Circle" /></td>
<td>a blue circle</td>
</tr>
<tr>
<td><img src="image2.png" alt="Circle" /></td>
<td>the circle size</td>
</tr>
<tr>
<td><img src="image3.png" alt="Arrow" /></td>
<td>an arrow</td>
</tr>
<tr>
<td><img src="image4.png" alt="Arrow" /></td>
<td>the width of the arrows</td>
</tr>
<tr>
<td><img src="image5.png" alt="Arrow" /></td>
<td>the color of the arrows</td>
</tr>
</tbody>
</table>

You can display the information about a rule (KI, Lift, Support...) by pointing a rule with the cursor.
4.3.3.4.1 Customizing the Graph

You can customize the width and/or the color of the arrows connecting the itemsets.

To Customize the Rules Display:
1. Click the *Customize Rules Display* button.
   
   A dialog opens allowing you to set the graphics visual parameters.

To Customize the Arrows Width:
2. In the *Edge Width list*, select the metrics you want to use.

To Customize the Arrows Color:
3. In the *Edge Color list*, select the metrics you want to use.
   
   The color varies from green to red, green being the lowest value and red the highest value.

To Display the Edge Width Value:
4. Check the *Display Edge Width Value as Label* option.
   
   The corresponding value is displayed for each arrow.
The toolbar located above the graph offers the following options:

<table>
<thead>
<tr>
<th>Icon</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Image](91x702 to 103x714)</td>
<td>Save as a png image</td>
</tr>
<tr>
<td>![Image](91x679 to 102x691)</td>
<td>Customize the graph</td>
</tr>
<tr>
<td>![Image](91x656 to 103x667)</td>
<td>Fit the graph to the screen size</td>
</tr>
<tr>
<td>![Image](91x633 to 135x644)</td>
<td>Zoom in or out</td>
</tr>
<tr>
<td>![Image](91x610 to 103x621)</td>
<td>Clear the current graph</td>
</tr>
</tbody>
</table>

### 4.3.3.4.2 Statistical Reports

To help you analyze your modeling results and possibly share these results with your colleagues, managers, partners or customers, Association Rules provides you with a set of statistical reports that can be saved in a variety of formats (html, pdf, rtf).

The following table presents the four types of tables available to you.

<table>
<thead>
<tr>
<th>The screen …</th>
<th>Presents the following information …</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rules Details (Support in Percentage)</strong></td>
<td>The association rules are displayed in the form: <em>item_1&amp;item_2&amp;...&amp;item_n-&gt;item_consequent</em>. The following information are given for each rule: confidence, KI, RuleSupport, ConsequentSupport, AntecedentSupport.</td>
</tr>
<tr>
<td><strong>Rules Details (Support in Sessions Count)</strong></td>
<td>This presents the same table as the previous one, but the support is displayed in number of sessions.</td>
</tr>
<tr>
<td><strong>Items Statistics</strong></td>
<td>This presents the list of the items treated by the model. The items frequency is given with respect to the transactions and not the sessions. For instance, if one item is present n times in one session, it will be counted n times and its percentage will be computed on the total count of the transactions.</td>
</tr>
</tbody>
</table>

### 4.4 Scenario 2: Modeling Using Sequence Mode

#### Presentation

Usually, the transactions have a temporal dimension, meaning that transactions occur one after another. In this case, the use of a standard Association Rules generation engine is not sufficient since it does not take into account the sequence of the transactions. To make the most of this temporal dimension, the association rules for which the antecedent occurs before the consequent need to be identified: this is the notion of Sequence.
Summary of the Modeling Settings to Use

In this step, you will follow the default scenario without using the Sequence mode (see "Scenario 1: Standard Modeling with Association Rules").

The following table summarizes the modeling settings to be used. It should be sufficient for users who are already familiar with SAP BusinessObjects Predictive Analytics.

<table>
<thead>
<tr>
<th>Task(s)</th>
<th>Screen</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifying the data source</td>
<td>Select a Data Source</td>
<td>In the field <em>Data Set</em>, select the file <code>website_references.csv</code></td>
</tr>
<tr>
<td>Selecting a cutting strategy</td>
<td>Cutting Strategy</td>
<td>Cutting strategy: random</td>
</tr>
<tr>
<td>Specifying a description file</td>
<td>Data Description</td>
<td>In the field <em>Description</em>, select the file <code>website_references_desc.csv</code></td>
</tr>
<tr>
<td>Specifying an Events file</td>
<td>Events Data Source</td>
<td>In the field <em>Events</em>, select the file <code>website_transactions.csv</code></td>
</tr>
<tr>
<td>Specifying a description file</td>
<td>Events Data Description</td>
<td>In the field <em>Description</em>, select the file <code>website_transactions_desc.csv</code></td>
</tr>
<tr>
<td>Activate the Sequence mode</td>
<td>Association Rules Parameters Settings</td>
<td>• Check the field <em>Activate Sequence Mode</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• In the field <em>Date</em>, select the variable <em>Time</em>.</td>
</tr>
</tbody>
</table>

4.4.1 Step 2 - Defining the Modeling Parameters

4.4.1.1 Setting Association Rules Parameters

The screen *Association Rules Extraction Parameters* enables you to set Association Rules parameters by:

• Joining your reference data with your transaction data.
• Setting the Specific Modeling Parameters.

To set the parameters:

1. On the *Association Rules Extraction Parameters* screen, select the join column for both the log and reference data sets.
   
   For our example, select the column `<SessionID>` in both data sets.

2. Select the *Item Column*.
   
   For our example, select the column `<Page>`.

3. Set the *Minimum Support*.
   
   For our example, enter `<1>`.
4. Check the option *Activate Sequence Mode*.
5. Select Time as the *Date column*.
6. Click *Next*.

### 4.4.1.2 Understanding Association Rules Parameters

This section describes only the *Sequence Mode*. The parameters Minimum Support, Minimum Confidence and Maximum Length are described in the section Understanding Association Rules Parameters of Scenario 1.

**Sequence Mode**

When you activate the *Sequence mode*, *Association Rules* computes specific information relative to this notion. The following parameters are available in the contextual menu *Select the Columns to Display* on the panel *Association Rule Exploration*:

- the *Sequence Support* indicates the number of sessions that verify the rules and in which the antecedent occurs before the consequent.
- the *Sequence Confidence* indicates the percentage of sessions verifying the consequent among those verifying the antecedent but only for the antecedents occurring before their consequent.
- the *Sequence KI* is only measured on the rules for which the antecedent occurs before the consequent.
- the *Sequence Lift* is only measured on the rules for which the antecedent occurs before the consequent.
- the *Sequence Ratio* measures the percentage of sessions where the antecedent occurs before the consequent in relation to all the sessions verifying the rule.

Also additional temporal measures will be computed:

- *Minimum Duration* indicates the minimum amount of time observed between an antecedent and its consequent. This value is expressed in seconds if the date is in a date or a datetime format.
- *Maximum Duration* indicates the maximum amount of time observed between an antecedent and its consequent. This value is expressed in seconds if the date is in a date or a datetime format.
- *Average Duration* indicates the average amount of time observed between an antecedent and its consequent. This value is expressed in seconds if the date is in a date or a datetime format.

Finally, only the Association Rules for which the *Sequence Support* is superior to the *Minimum Support* and the *Sequence Confidence* is superior to the *Minimum Confidence* will be generated.

**Note**

Computing and Filtering Association Rules taking into account the temporal dimension requires an additional sweep on the transaction file.

### Understanding Advanced Parameters

The Advanced Parameters are described in Scenario 1 (see "Understanding Advanced Parameters").
4.4.2 Step 3 - Generating the Model

Once the modeling parameters are defined, you can generate the model.

To generate the model, on the screen Summary of Modeling Parameters, click the Generate button. The screen Training the Model will appear. The model is being generated. A progress bar allows you to follow the process.

There is two ways for you to follow the progress of the generation process:

- The Progress Bar displays the progression for each step of the process. It is the screen displayed by default.
- The Detailed Log displays the details of each step of the process.

To Display the Detailed Log:
1. Click the Show Detailed Log button.
   The log will be displayed.

To display the Progression Bar:
2. Click the Show Progression button.
   The progression bar screen appears.

To Stop the Learning Process:
3. Click the Stop Learning Process button.
4. Click Previous.
   The Summary of Modeling Parameters screen appears.
5. Go back to the Checking Modeling Parameters section.

i Note

At the end of the generation process, a summary of the model results will appear.

For more information on the model summary, go to step 4 (see "Step 4 - Analyzing and Understanding the Generated Model").

4.4.3 Step 4: Model Debriefing

Automated Analytics allows you to analyze and understand the model generated thanks to a suite of plotting tools describing:

- all the items treated by the model,
- the rules generated by the model.

To Display the Model Overview:
1. On the Using the Model menu, select the Model Overview option.
2. Click Next. The Model Overview screen will appear.

i Note

If you have built more than one model in the same session, all model debriefing will be displayed on this screen sorted by Date of Build.
4.4.3.1 Understanding the Model Debriefing

The screen Model Debriefing is the same as described in Scenario 1 (see "Understanding the Model Debriefing"), except for two added parameters corresponding to the Sequence Mode.

The following table details the additional information provided in section Parameters:

<table>
<thead>
<tr>
<th>Name</th>
<th>Significance</th>
<th>For example...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence Mode</td>
<td>indicates if the Sequence Mode has been activated.</td>
<td>Activated</td>
</tr>
<tr>
<td>Date Variable</td>
<td>indicates which variable has been used as Date Variable for the sequences.</td>
<td>Time</td>
</tr>
</tbody>
</table>

Association Rules Exploration

The panel Association Rules Exploration is the same as described in Scenario 1, except for added filters corresponding to the Sequence Mode:

- **Sequence Support**, that is the number of sessions in which the antecedent occurs before the consequent.
- **Sequence Ratios**, that is the percentage of session where the consequent occurs after the antecedent for a specified rule.
- **Sequence Confidence**, that is the rule confidence in the sequence mode.
- **Sequence KI**, that is the rule KI in the sequence mode.
- **Sequence Lift**, that is the rule Lift in the sequence mode.
- **Minimum Duration**, that is the minimum amount of time observed between an antecedent and its consequent. This value is expressed in seconds if the date is in a date or a datetime format.
- **Maximum Duration**, that is the maximum amount of time observed between an antecedent and its consequent. This value is expressed in seconds if the date is in a date or a datetime format.
- **Average Duration**, that is the average amount of time observed between an antecedent and its consequent. This value is expressed in seconds if the date is in a date or a datetime format.

Debriefing Tables

For details on the Debriefing Tables, go to the section Understanding the Model Debriefing of Scenario 1.

Model Parameters

For details on the Model Parameters, go to the section Understanding Association Rules Parameters of Scenario 1.
4.4.4 Step 5: Using the Model

4.4.4.1 Saving the Model

Once a model has been generated, you can save it. Saving it preserves all the information that pertains to that model, that is, the modeling parameters, the rules generated, and so on.

1. On the Using the Model screen, click Save the Current Model.

   The Saving the Model screen opens.

2. Above the Browse button, select the type of file to create.

3. Complete the following fields:
   - **Model Name**: this field allows you to associate a name with the model. This name will then appear in the list of models to be offered when you open an existing model.
   - **Description**: this field allows you to enter the information of your choosing, such as the name of the training data set used. This information will help you identify your model for a later use.
   - **Folder**: depending on which option you selected, this field allows you to specify the ODBC source, the memory store or the folder in which you want to save the model.
   - **File/Table**: this field allows you to enter the name of the file or table that will contain the model. If you have selected the option Text File, the name of the file must contain one of the following format extensions .txt (text file in which the data is separated by tabs) or .csv (text file in which the data is separated by commas).

4.4.4.2 Applying a Model

The currently open model may be applied to additional data sets. The model allows you to perform predictions using the application data sets, and specifically, to make recommendations based on the data sets used to generate the model.

To Apply the Model to a New Data Set:

1. On the Using the Model panel, select the Apply Model option.

2. Click Next.

   The Applying the Model screen is displayed.

3. In the Application Data Set section, click the Browse button to select:
   - In the **Folder** field, the folder which contains your data set,
   - In the **Data** field, the name of the file corresponding to your data set.

4. In the Results Generated by the Model section, enter:
   - In the **Folder** field, the name of the folder where you want to save the output of the apply,
   - In the **Data** field, the name of the file containing the output of the apply.

5. In the Generate list, select the type of output you want to generate:
   - **Predicted Value Only**: generates basic outputs, that is, the session key, the ID of the rule used to find the consequent and the consequent itself.
   - **Consequents Optimized by KI**: generates basic outputs. If more than one rule give the same consequent for a session, the rule presenting the best KI (predictive power) will be selected.
○ **Consequents Optimized by Confidence**: generates basic outputs. If more than one rule give the same consequent for a session, the rule presenting the best Confidence will be selected.

○ **Predicted Value with Full Rules Description**: generates the extended outputs, that is, the session key, the rule ID, the consequent, the antecedent, the KI, the confidence and the rule support.

○ **Consequents Optimized by KI with Full Rules Description**: generates the extended outputs. If more than one rule give the same consequent for a session, the rule presenting the best KI will be selected.

○ **Consequents Optimized by Confidence with Full Rules Description**: generates the extended outputs. If more than one rule give the same consequent for a session, the rule presenting the best Confidence will be selected.

6. Click **Next**. The **Apply Events Data** screen is displayed.

7. Click the **Browse** button to select:
   ○ In the **Folder** field, the folder which contains your events data set.
   ○ In the **Data** field, the name of the file corresponding to your events data set.

8. Click **Next**.

   A dialog box is displayed, asking if you want to replace the current events data set by the one you just selected.

9. Click **Apply**.

   A progress bar is displayed allowing you to follow the apply process.

10. At the end of the applying process, click **Next** to go back to the **Using the Model** screen.

---

### 4.4.4.2.1 Constraints of Model Use

In order to apply a model to a data set, it must contain the following variables:

- all the variables with the "input" role, meaning all variables that have not been excluded in the variable selection step (even if all are not used by the model)

  **Caution**

  Even if an auto selection has been used, all the initial input variables are mandatory.

- all model keys even if they have been excluded (except the KxIndex generated by Automated Analytics).

Also the data set generated as a result of the application must contain the following variables:

- the estimator variable (prefixed rr_) if the **Predicted** value is requested in the **Advanced Apply Settings** (which is the default behavior).
- the target variables, to retrieve their values if they exist in the application data set,
- all model keys.

To make sure that the data sets are consistent, use the mapping feature provided in the panel **Applying the Model**.
4.4.4.2.2 Deactivating Association Rules

Not all the association rules have the same importance when applying: some of them are relevant whereas other ones do not give any information. That is why Association Rules allows you to deactivate association rules before applying them.

1. Before applying the model, go to the Association Rules Exploration panel. For a complete description of this panel, see section Association Rules Exploration.
2. Set a filter.
3. Click the Search button to display a set of association rules in the Association Rules Found section.
4. Right-click the association rules list. A contextual menu is displayed.

The first section of the menu, above the separator, allows you to change the application state of the association rules:

<table>
<thead>
<tr>
<th>The Option...</th>
<th>Allows You to...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activate the Selected rule(s)</td>
<td>activate the selected rules for the next application</td>
</tr>
<tr>
<td>Deactivate the selected rules</td>
<td>deactivate the selected rules for the next application</td>
</tr>
<tr>
<td>Activate All</td>
<td>activate all association rules found by Association Rules</td>
</tr>
<tr>
<td>Deactivate All</td>
<td>deactivate all association rules found by Association Rules</td>
</tr>
</tbody>
</table>

The second section of the menu, below the separator, allows you to control the association rules displayed in the table:

- Display all the rules is the default display. All the association rules are displayed; the activated ones as well as the deactivated ones.
- Display only the activated rules
- Display on the Deactivated rules

The deactivated association rules will not be used in the next application of the model.

4.4.4.3 Opening a Model

Once saved, models may be opened and reused in the application.

1. On the main screen, click on Load a Model.
   The Opening a Model screen will appear.
2. Select the file type depending on the format of the model you want to open.
3. Click Browse.
4. Select the folder that holds the model that you want to open. The list of models contained in that folder will appear.

The following table lists the information provided for each model.
<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name under which the model has been saved</td>
<td>Character string</td>
</tr>
<tr>
<td>Class</td>
<td>Class of the model, that is the type of the model</td>
<td>○ Kxen.Classification: Classification/Regression with nominal target</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ Kxen.Regression: Classification/Regression with continuous target</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ Kxen.Segmentation: Clustering with SQL Mode</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ Kxen.Clustering: Clustering without SQL Mode</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ Kxen.TimeSeries: Time Series</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ Kxen.AssociationRules: Association Rules</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ Kxen.SimpleModel: Classification/Regression and Clustering multi-target models, any other model</td>
</tr>
<tr>
<td>Version</td>
<td>Number of the model version when the model has been saved several times</td>
<td>Integer starting at 1</td>
</tr>
<tr>
<td>Date</td>
<td>Date when the model has been saved</td>
<td>Date and time in the format yyyy-mm-dd hh:mm:ss</td>
</tr>
<tr>
<td>Comment</td>
<td>Optional user defined comment that can be used to identify the model</td>
<td>Character string</td>
</tr>
</tbody>
</table>

5. Select a model from the list
6. Click Open.

The Using the Model screen will appear.
Creating the Association Rules Model

The following code describes the method to create an Association Rules model. A default model is created including an association rules transform.

```
#creating a simple model with an association rules transform
createModel KXEN.SimpleModel myModel
myModel.pushTransformInProtocol Default KXEN.ASSOCIATION RULES myASSOCIATION RULES
```

Setting the Simple Model Parameters

The cutting strategy is the only model parameter that has to be set.

```
#setting the model’s parameters
myModel.getParameter ""
myModel.changeParameter Parameters/CutTrainingPolicy random
myModel.validateParameter
```

Opening the Reference and the Events Data Set

In this step, the Reference data set website_references.csv is opened with its description website_references_desc.csv. The Events data set website_transactions.csv, containing the transactions, is also opened with its description website_transactions_desc.csv.

```
#opening the data sets
myModel.openNewStore KXEN.FileStore C:\Program Files\KXEN\KxenComp\Samples \ASSOCIATION RULES
myModel.newDataSet Training website_references.csv
myModel.readSpaceDescription Training website_references_desc.csv
myModel.newDataSet Events website_transactions.csv
myModel.readSpaceDescription Events website_transactions_desc.csv
```

Setting the Association Rules Transform Parameters

The association rules transform parameters are the following:
• the Transaction parameters, which are used to set the information relative to the Events data source,
• the Reference parameters, which are used to set the information relative to the Reference data source,
• the Association Rules engine parameters. The algorithm used to generate the association rules is Frequent Pattern Vertical (FPV), this is why the Association Rules parameters are located under a branch called FPV.
• ExtraMode [read-write] is a special flag allowing to set the type of outputs that Association Rules will generate. It can take the following six values:
  ○ No Extra: generates basic outputs, that is the session key, the ID of the rule used to find the consequent and the consequent itself.
  ○ Optimized by KI: generates basic outputs. If more than one rule give the same consequent for a session, the rule presenting the best KI will be selected.
  ○ Optimized by Confidence: generates basic outputs. If more than one rule give the same consequent for a session, the rule presenting the best Confidence will be selected.
  ○ Full Description: generates the extended outputs, that is the session key, the rule ID, the consequent, the antecedent, the KI, the confidence and the rule support.
  ○ Full Description and Optimized by Confidence: generates the extended outputs. If more than one rule give the same consequent for a session, the rule presenting the best Confidence will be selected.
  ○ Full Description and Optimized by KI: generates the extended outputs. If more than one rule give the same consequent for a session, the rule presenting the best KI will be selected.
• DateColumnName [Sequence mode only] is the column in which the date is stored.
• SequencesMode [Sequence mode only] is a flag specifying if the Sequence mode of Association Rules is activated. True (or 1) means that the Sequence mode is activated, False (or 0) means that the Sequence mode is deactivated.

Transactions Section

• Transactions/LogSpaceName indicates Automated Analytics role set for the Transactions space name.
• Transactions/TIDColumnName indicates the name of the transactions key variable.
• Transactions/ItemIDColumnName indicates the name of the item variable.

References Section

References/TIDColumnName indicates the name of the reference key variable.

Association Rules Engine

The Association Rules engine can be used to generate general association rules or specific association rules to replace the missing values. This second section can only be used by other features.
Default Use

ARulesEngineParameters/FPV/Activated indicates whether the FPV algorithm is activated or not. This value should always be set to true (default = true).

ARulesEngineParameters/FPV/MinimumSupport gives the minimum support required for a rule. With a value > 1, we consider the number of sessions. With a value between 0 and 1, we consider a percentage of the number of sessions (default = 1, required value > 0).

ARulesEngineParameters/FPV/MinimumConfidence gives the minimum threshold for the confidence of a rule (default = 0.5, required a value between 0 and 1).

ARulesEngineParameters/FPV/MaxLength saves the maximum size of a rule. A rule has a minimum size equal to 2: the antecedent plus the consequent (default = 4, required value >= 2).

ARulesEngineParameters/FPV/SearchMethod indicates the search method the FPV has to use. When the value equals 1, the Association Rules engine uses the basic FPV algorithm. When the value equals 2, the Association Rules uses a modified FPV algorithm: the rules are generated regarding the frequency order of their consequent item and the memory used is limited (default = 2, required value equals 1 or 2).

ARulesEngineParameters/FPV/SkipDerivedRules indicates to FPV to generate only the primary rules and thus skip the derived rules. For instance, given R1: X => Z and R2: X & Y => Z two association rules, R1 is called a primary rule and R2 a derived rule. The option SkipDerivedRules is only available with a SearchMethod parameter equal to 2 (default = true).

ARulesEngineParameters/FPV/SkipReducibleItemsets indicates if FPV has to generate the reducible itemsets (default = false). An itemset is non reducible if one of its items cannot be removed without changing the transaction space.

ARulesEngineParameters/FPV/TidSet indicates the storage method of the transactions associated to an itemset (default = 2, required a value equal to 1 or 2).

ARulesEngineParameters/FPV/ChunkSize saves the size of the chunks (in number of sessions) use by FPV to import and generate the rules. With a value equal to 0, the chunk strategy will not be used and ALL the sessions will be imported before generating the rules (default = 0, required value >= 0).

Description of the Related Script

```plaintext
#setting the transform parameters
myASSOCIATION RULES.getParameter ""
myASSOCIATION RULES.changeParameter "Parameters/Transactions/LogSpaceName" Events
myASSOCIATION RULES.changeParameter "Parameters/Transactions/TIDColumnName" SessionID
myASSOCIATION RULES.changeParameter "Parameters/Transactions/ItemIDColumnName" Page
myASSOCIATION RULES.changeParameter "Parameters/References/TIDColumnName" SessionID
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/Activated" true
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/MinimumSupport" 0.01
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/MinimumConfidence" 0.5
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/MaxLength" 4
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/SkipDerivedRules" true
myASSOCIATION RULES.validateParameter
```
Specify Filters on the Association Rules

This section is used to specify filters on the items that should or should not appear as consequents in the generated association rules.

ARulesEngineParameters/RulesGenerationFilters/ConsequentFilters/IncludedList contains the list of the items that should appear as consequents.

ARulesEngineParameters/RulesGenerationFilters/ConsequentFilters/ExcludedList contains the list of the items that should not appear as consequents.

The list of the included items is taken into account before the list of the excluded items. The list of the excluded items is used if the list of the included items is empty.

Description of the Related Script

```plaintext
#set the list of the consequents to keep
myInfiniteInsight™ Modeler - Association Rules.getParameter ""
myInfiniteInsight™ Modeler - Association Rules.bindParameter "Parameters/ARulesEngineParameters/RulesGenerationFilters/ConsequentFilters/IncludedList"
ConsParam.insert "//HM_Body.tmpl" fakeParam_0
myInfiniteInsight™ Modeler - Association Rules.validateParameter
delete ConsParam
```

Missing Values Specific Additional Parameters

ARulesEngineParameters/FPV/DefaultTargetEnable indicates whether to generate the rules that have the most frequent item of a group as the consequent part. This parameter should be set to false, because this parameter is useless to deduce the rules used to fill the missing values (default = false).

ARulesEngineParameters/FPV/UseMissInfo indicates that the confidence of the rules will be computed taking into account the missing information for each variable. This parameter is only used when the Association Rules engine is used to replace the missing values.

Activate the Sequence Mode

```plaintext
#Activating the Sequence Mode
myKAR.getParameter ""
myKAR.changeParameter "Parameters/SequencesMode" true
myKAR.changeParameter "Parameters/DateColumnName" Time
myKAR.validateParameter
```
### 5.1 Learning the Model

Once all the parameters have been set, the learning step is launched.

```plaintext
#learning the model
myModel.sendMode learn
```

### 5.2 Applying the Model

Once the model has been generated, you can apply it. Before applying the model you can change the value of the ExtraMode parameter (see above in section Setting the Association Rules Transform Parameters).

```plaintext
# apply the model with default ExtraMode
mymodel.newDataSet ApplyIn website_references.txt
mymodel.openNewStore Kxen.FileStore Saved
mymodel.newDataSet ApplyOut out_default_website.csv
mymodel.sendMode apply
```

Several association rules can be deactivated before applying the model because they are not relevant. There are two ways to realize that:

- Specify a list of association rule identifiers to deactivate in the path `Parameters/ARulesEngineParameters/ApplyActivationOptions/RulesExcludedList`

- Specify a list of interest items, that are items for which all the association rules having one of these items as a consequent should be activated, in the path `Parameters/ARulesEngineParameters/ApplyActivationOptions/ActivatedConsequentsList`. The Association Rules that have a consequent not included in this list will not be activated.

**Note**

The Association Rules having their consequent item included in the interest items list but having their identifier registered in the excluded list, these association rules will be excluded for the next application.

### Related Script

```plaintext
# deactivate several rules before applying again
# First specify a list of association rules to activate by specifying the interest
# consequent items
myKAR.getParameter ""
myKAR.bindParameter "Parameters/ARulesEngineParameters/ApplyActivationOptions/ActivatedConsequentsList" ConsParam
ConsParam.insert "/shop/order1.tmpl" fakeParam_0
ConsParam.insert "/shop/order2.tmpl" fakeParam_1
ConsParam.insert "/shop/order3.tmpl" fakeParam_2
ConsParam.insert "/shop/order4.tmpl" fakeParam_3
```
5.3 Saving the Model

Once a model has been generated, you can save it. Saving it preserves all the information that pertains to that model that is the modeling parameters, the association rules and so on.

Note

The directory, in which the model is saved, must preexist.

```
# saving the model in a Saved directory
myModel.openNewStore KXEN.FileStore Saved
myModel.saveModel ASSOCIATION RULES.txt "ASSOCIATION RULES on website_references with <Page> as the item variable."
```

5.4 Opening an Existing Model

Once saved, models may be opened and reused in SAP BusinessObjects Predictive Analytics.

For this scenario:

```
# restoring the last saved model
createStore KXEN.FileStore myRestoreStore
myRestoreStore.openStore "Saved"
setDefaultUserPassword "" ""
myRestoreStore.restoreLastModelID myModel
```

5.5 Summarizing the KxShell script

Below is an example of a complete script you can build.

```
# creating a simple model with an association rules transform
createModel KXEN.SimpleModel myModel
myModel.pushTransformInProtocol Default KXEN.ASSOCIATION RULES myASSOCIATION RULES
```
#setting the model parameters
myModel.getParameter ""
myModel.changeParameter Parameters/CutTrainingPolicy random
myModel.validateParameter

#opening the data sets
myModel.openNewStore KXEN.FileStore C:\Program
Files\KXEN\KxenCompV3\Samples\ASSOCIATION RULES
myModel.newDataSet Training website_references.csv
myModel.readSpaceDescription Training
website_references_desc.csv
myModel.newDataSet Events website_transactions.csv
myModel.readSpaceDescription Events
website_transactions_desc.csv

#setting the transform parameters
myASSOCIATION RULES.getParameter ""
myASSOCIATION RULES.changeParameter "Parameters/Transactions/LogSpaceName" Events
myASSOCIATION RULES.changeParameter "Parameters/Transactions/TIDColumnName" SessionID
myASSOCIATION RULES.changeParameter "Parameters/Transactions/ItemIDColumnName" Page
myASSOCIATION RULES.changeParameter "Parameters/References/TIDColumnName" SessionID
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/Activated" true
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/MinimumSupport" 0.01
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/MinimumConfidence" 0.5
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/MaxLength" 4
myASSOCIATION RULES.changeParameter "Parameters/ARulesEngineParameters/FPV/SkipDerivedRules" true
myASSOCIATION RULES.validateParameter

#Activating the Sequence Mode
myASSOCIATION RULES.getParameter ""
myASSOCIATION RULES.changeParameter "Parameters/SequencesMode" true
myASSOCIATION RULES.changeParameter "Parameters/DateColumnName" Time
myASSOCIATION RULES.validateParameter

#learning the model
myModel.sendMode learn

#APPLY the association rules
#apply the model with default ExtraMode
myModel.newDataSet ApplyIn website_references.txt
myModel.openNewStore Kxen.FileStore Saved
myModel.newDataSet ApplyOut out_default_website.csv
myModel.sendMode apply

#apply only some of the association rules
#deactivate several rules before applying again
#First specify a list of association rules to activate by specifying the interest
# consequent items
myASSOCIATION RULES.getParameter ""
myASSOCIATION RULES.bindParameter "Parameters/ARulesEngineParameters/
ApplyActivationOptions/ActivatedConsequentsList" ConsParam
ConsParam.insert "/shop/order1.tmpl" fakeParam_0
ConsParam.insert "/shop/order2.tmpl" fakeParam_1
ConsParam.insert "/shop/order3.tmpl" fakeParam_2
ConsParam.insert "/shop/order4.tmpl" fakeParam_3
ConsParam.insert "/shop/order5.tmpl" fakeParam_4

#Then, specify a list of association rules to exclude by specifying their identifiers
myASSOCIATION RULES.bindParameter "Parameters/ARulesEngineParameters/
ApplyActivationOptions/RulesExcludedList" RulesParam
RulesParam.insert "Rule850" fakeParam_5
RulesParam.insert "FakeRule" fakeParam_7
myASSOCIATION RULES.validateParameter
delete ConsParam
#applying with the new options
mymodel.newDataSet ApplyIn website_references.txt
mymodel.openNewStore Kxen.FileStore Saved
mymodel.newDataSet ApplyOut out_OrdersAsConsequent ExceptRule850_website.csv
mymodel.sendMode apply
#saving the model in a Saved directory
myModel.openNewStore KXEN.FileStore Saved
myModel.saveModel ASSOCIATION RULES.txt "ASSOCIATION RULES on website_references
with
<Page> as the item variable."
Important Disclaimers and Legal Information

Coding Samples

Any software coding and/or code lines / strings ("Code") included in this documentation are only examples and are not intended to be used in a productive system environment. The Code is only intended to better explain and visualize the syntax and phrasing rules of certain coding. SAP does not warrant the correctness and completeness of the Code given herein, and SAP shall not be liable for errors or damages caused by the usage of the Code, unless damages were caused by SAP intentionally or by SAP's gross negligence.

Accessibility

The information contained in the SAP documentation represents SAP’s current view of accessibility criteria as of the date of publication; it is in no way intended to be a binding guideline on how to ensure accessibility of software products. SAP in particular disclaims any liability in relation to this document. This disclaimer, however, does not apply in cases of willful misconduct or gross negligence of SAP. Furthermore, this document does not result in any direct or indirect contractual obligations of SAP.

Gender-Neutral Language

As far as possible, SAP documentation is gender neutral. Depending on the context, the reader is addressed directly with "you", or a gender-neutral noun (such as “sales person” or “working days”) is used. If when referring to members of both sexes, however, the third-person singular cannot be avoided or a gender-neutral noun does not exist, SAP reserves the right to use the masculine form of the noun and pronoun. This is to ensure that the documentation remains comprehensible.

Internet Hyperlinks

The SAP documentation may contain hyperlinks to the Internet. These hyperlinks are intended to serve as a hint about where to find related information. SAP does not warrant the availability and correctness of this related information or the ability of this information to serve a particular purpose. SAP shall not be liable for any damages caused by the use of related information unless damages have been caused by SAP’s gross negligence or willful misconduct. All links are categorized for transparency (see: http://help.sap.com/disclaimer).