IBM® Analytical Decision Management
Version 17 Release 0

IBM® SPSS® Rules Management

IBM

Product Information

This edition applies to version 17, release 0, modification 0 of IBM Analytical Decision Management and to all subsequent releases and modifications until otherwise indicated in new editions.

Contents

Preface

About IBM Business Analytics

IBM Business Analytics software delivers complete, consistent, and accurate information that decision-makers trust to improve business performance. A comprehensive portfolio of business intelligence, predictive analytics, financial performance and strategy management, and analytic applications provides clear, immediate and actionable insights into current performance and the ability to predict future outcomes. Combined with rich industry solutions, proven practices, and professional services, organizations of every size can drive the highest productivity, confidently automate decisions, and deliver better results.

As part of this portfolio, IBM SPSS Predictive Analytics software helps organizations predict future events and proactively act upon that insight to drive better business outcomes. Commercial, government, and academic customers worldwide rely on IBM® SPSS® technology as a competitive advantage in attracting, retaining, and growing customers, while reducing fraud and mitigating risk. By incorporationg IBM SPSS software into their daily operations, organizations become predictive enterprises - able to direct and automate decisions to meet business goals and achieve measurable competitive advantage. For further information or to reach a representative, visit http://www.ibm.com/spss.

Technical support

Technical support is available to maintenance customers. Customers may contact Technical Support for assistance in using IBM Corp. products or for installation help for one of the supported hardware environments. To reach Technical Support, see the IBM Corp. web site at http://www.ibm.com/support. Be prepared to identify yourself, your organization, and your support agreement when requesting assistance.

Chapter 1. About IBM Analytical Decision Management

Product overview

IBM Analytical Decision Management brings the benefit of predictive analytics to real business problems, allowing you to build custom applications tailored to your customers or industry. While applications are typically configured to solve very specific problems, all are based on a common set of capabilities:

- Automate decisions using business rules.
- Add insight using predictive models.
- Use prioritization, optimization, or simulation to reach the best decision based on the above.

A number of packaged applications are available, tailored to solving specific business problems. Contact your sales representative for more information.

IBM SPSS Rules Management

IBM SPSS Rules Management provides a central tool for creating and editing shared rules, which can be used throughout applications to select and process records, and to automate decisions accordingly. While rules can also be created and saved within each application, IBM SPSS Rules Management allows you to save rules as separate objects that can be referenced from multiple applications. For example, a rule to exclude customers who are below legal age could be shared across all applications to enforce a global policy, and any changes to the shared rule would be applied across all applications.

Note that IBM SPSS Rules Management may not be installed at all sites or available to all users. As an alternative, rules can also be defined and saved locally for use within each application.

What's new in version 17

IBM Analytical Decision Management 17 (17.0.0.0) adds the following new features:

Following is a description of the main new features:

Expression Builder. The Expression Builder has been enhanced to allow more complex and powerful expressions. The Expression Builder in IBM Analytical Decision Management now better matches the Expression Builder in IBM SPSS Modeler client. See http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_expressions.dita.

Derived tab. In most IBM Analytical Decision Management applications, you can add additional fields to your project data model such as expressions, model output, fields created from segment rules, or fields from secondary tables. The new **Derived** sub tab of the Data tab displays any fields that are to be included as part of the data. See http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_app_derived.dita.

Secondary Tables section. A new Secondary Tables section has been added to most IBM Analytical Decision Management applications. This replaces the various **Add/Edit additional tables** links found in previous releases. See http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/dm_app_additional_fields.dita.

Multi-factor authentication. Functionality that supports multi-factor authentication (MFA) has been added to IBM SPSS Collaboration and Deployment Services 7, IBM SPSS Modeler 17, and IBM Analytical Decision Management 17. These enhancements can improve security for large-scale enterprise environments. For example, SPSS products can be integrated with a company's existing Kerberos Single Sign On infrastructure and with SmartCard infrastructures (when a company's employees use security

badges and card readers for machine and application authentication, for example). See
http://www-01.ibm.com/support/knowledgecenter/SS6A3P_17.0.0/configurableapps/
dms_install_security.dita.

Data source mapping. When possible, IBM Analytical Decision Management will now automatically map compatible storage types when a user adds an additional data source. If the type is not compatible, the user will be allowed to convert types.

File storage types. Users can now set the storage type and the format when reading in fixed and variable format data files, to ensure that values are read correctly. For example, if data source type is **File** and the file format is **Variable length fields in plain text** when adding a new data source in IBM Analytical Decision Management, the Data Source dialog will now have a new **Specify Data Format** section that allows the user to specify the input type and the format of each field. This is similar to the functionality available in IBM SPSS Modeler client.

Project summary. For applications with a Deploy tab, a new **Project Summary** section has been added. Use this section to validate the project and display important project summary information, including any errors found that may impact deployment. The project summary displays information such as all attribute references and all repository references, and can be exported as a *.csv file. You can also move to other tabs of the application to fix errors and then return to the project summary on the Deploy tab.

Model references. When you reference models from within an IBM Analytical Decision Management application, the behavior of the model link has changed. When you click a model reference link, the model viewer dialog opens, allowing you to view details about the model. Or you can now right-click the model reference to select the option **Reference** or **Build**. When you click a rule reference, the rule viewer will open (as in previous releases). Or you can right-click a rule reference and select **Reference**.

Analytic Server credential suffix. Administrators can use browser-based IBM SPSS Collaboration and Deployment Services Deployment Manager to configure a suffix to append to Analytic Server credentials. This will derive the name of the credential that will be used for Analytic Server connections in IBM Analytical Decision Management.

IBM Cognos credentails. When importing IBM Cognos BI database data or single list reports for use in IBM Analytical Decision Management, you can now select an IBM Cognos credential stored in the IBM SPSS Collaboration and Deployment Services Repository rather than entering user name and password information or using an anonymous connection.

Previous versions

Version 8
IBM SPSS Analytical Decision Management 8 extended users' ability to deploy analytics into everyday business processes, supporting a wider range of decisions with greater insight and control than ever before.
- Fine-tune and visualize the combined impact of rules, models, and optimization on business outcomes, with enhanced simulation capabilities.
- Solve a wider range of problems with new prebuilt applications and new optimization techniques such as price optimization and assortment optimization.
- Improved integration with IBM Operational Decision Management, and big data support via IBM SPSS Analytic Server integration.

Following is a description of the main new features that were added in version 8:

New installation technology. The product installation has been simplified and improved. It now uses Installation Manager technology. For more information, see the *IBM SPSS Analytical Decision Management Installation Guide*.

User interface enhancements. Many changes have been made to the user interface to improve usability.

IBM Analytical Decision Management for Predictive Maintenance. IBM SPSS Analytical Decision Management for Predictive Maintenance is a new prebuilt application that uses optimization to anticipate and prevent equipment failures before they occur. You can minimize costly downtime and make best use of scarce infrastructure and resources by predicting failures and scheduling the right correcting actions at the right time. You might anticipate which parts are likely to fail first and replace them at the appropriate time. Predicting when machines will fail helps to maximize profits and avoid losses from machines failing and being out of action.

IBM Analytical Decision Management for Demand Optimization. IBM SPSS Analytical Decision Management for Demand Optimization is a new prebuilt application that uses optimization to avoid costly overstocks and shortfalls by allocating the right amount of each product to each store. You can maximize profits and avoid losses due to inventory shortfalls and restocking fees by allocating the right amount of inventory to each outlet.

Simulation enhancements. New simulation capabilities combine the functionality of the previous **Simulate** button (formerly on the Define tab) with the previous **What If?** functionality (formerly on the Optimize, Prioritize, or Combine tab) into a single feature. The new and improved simulation functionality has its own toolbar in the upper-right corner of the user interface, and allows for scenarios to be saved with the project and persist across sessions.

Deployment and scoring. Various enhancements have been made to the Deploy/Score tab.

Application URL. IBM SPSS Collaboration and Deployment Services has added the ability to customize the URL for web-based applications, including the URL for launching IBM SPSS Analytical Decision Management. Contact your administrator for details. More information is available in the *Application context roots* chapter of the *IBM SPSS Collaboration and Deployment Services Repository Installation and Configuration Guide*.

Statistics file enhancements. IBM SPSS Statistics compressed, password-protected data files (*.sav* and *.zsav*) are now supported in IBM SPSS Analytical Decision Management.

Tree viewer enhancements When viewing individual model results in IBM SPSS Modeler Advantage, the name of each model is a link that opens a new dialog containing information about the model such as the target field and input fields. If the selected model is a *tree model*, then a tree viewer will also be displayed. The tree viewer has been enhanced. You can now perform such actions as zooming in and out of the tree, changing the tree orientation, navigating large trees, and expanding or collapsing individual nodes within the tree.

Advanced options on Optimize tab. For applications that use complex mathematical optimization (CPLEX), such as IBM SPSS Analytical Decision Management for Campaign Optimization, IBM SPSS Analytical Decision Management for Predictive Maintenance, and IBM SPSS Analytical Decision Management for Demand Optimization, advanced options are now available related to the CPLEX optimization algorithm. You can now export an Optimization Programming Language (OPL) model file and add custom settings.

Data View source. A new *Data View* node has been added to IBM SPSS Modeler 16 to include data in streams that was defined in an IBM SPSS Collaboration and Deployment Services **analytic data view**. The new Data View data source can also be used on the Data tab in IBM SPSS Analytical Decision Management. You can create a connection to use with IBM Operational Decision Management rules. For complex rules that use a hierarchical data model, note that administrators may restrict use of such rules because they can degrade performance.

Prior to version 8, external rules required flat input parameters that matched the IBM SPSS Analytical Decision Management project data model. But with this new analytic data view support, external rules

can now have hierarchical input parameters. See the topics "Data View source" on page 21 and "Locking data options" on page 53 for more information, and *Chapter 5* of the *Application Designer's Guide*.

IBM SPSS Analytic Server integration. In IBM SPSS Modeler 16, IBM SPSS Analytic Server has been introduced, allowing the use of Hadoop Distributed File System (HDFS) as a data source. IBM SPSS Modeler users can create streams using the HDFS Analytic Server source node and build models based on the Hadoop big data system. IBM SPSS Analytical Decision Management users can open these model streams in IBM SPSS Modeler Advantage for viewing (they cannot create new projects using Analytic Server data sources).

New chapter in Application Designer's Guide. A new chapter has been added to the *IBM SPSS Analytical Decision Management Application Designer's Guide*. It provides an example configuration of a custom optimization application. The standard IBM SPSS Analytical Decision Management for Campaign Optimization application uses a preconfigured Optimization Programming Language (OPL) model file called *CampaignOptimization.mod*. In some cases, the standard Campaign Optimization application and its OPL model might not suit the purposes of the business problem at hand. The new chapter walks through an example of a custom OPL model built using IBM ILOG CPLEX Optimization Studio, and a custom IBM SPSS Analytical Decision Management optimization application, for a fictional company.

Tabular evaluation results. A new **Include tabular results** option is available when running evaluations. Deselecting the option improves performance because it takes more time to run evaluations that include tabular results.

New setting for maximum number of nominal fields. A new setting has been added to the Properties dialog for setting the maximum number of members allowed for *Nominal* data fields. This is used on the Data tab. For more information, see "Properties" on page 14.

Single Sign On. ODBC connections can now use Single Sign On (SSO) for IBM SPSS Modeler Server and ODBC database connections with IBM SPSS Analytical Decision Management. Some configuration steps are required. See the topic Security considerations in the installation documentation.

SQL push-back. When the source data originates in a database, execution is pushed back to the database whenever possible (in a process known as "SQL push-back"). In general, IBM SPSS Analytical Decision Management supports the same level of SQL push-back as IBM SPSS Modeler but with the following limitations:

- If the project includes additional fields or selection rules which have references to models or external rules, then these references will not push back to the database, and this may inhibit the degree of push-back achievable for the project stream as a whole.
- In applications other than IBM SPSS Modeler Advantage, push-back is enabled only up to selection rules (global selections, and member selections on the first dimension). The degree of push-back may be increased if you choose an Integer field as the entity ID in the project data model.

Version 7.0.0.1

IBM SPSS Decision Management 7 Fix Pack 1 (7.0.0.1) added the following features:

Data Overview. A new data overview feature is available on the Data tab of all applications, on the Modeling tab and the Score tab in IBM SPSS Modeler Advantage, and on the Deploy tab of applications that include the *Score Now* feature. See the topic "Data overview" on page 24 for more information. This feature is only available with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). It isn't available with version 15.

Scoring Summary Report. After scoring a project, a new *Scoring Summary Report* section provides details such as where the data was scored to, file size (if applicable), database and table name (if applicable), model name or project name, and number of records scored, and in some cases a chart and/or table summarizing the model. The Scoring Summary Report section is available in IBM SPSS Modeler

Advantage, and on the Deploy tab of applications that include the Score Now feature. Note that the files size is only displayed if running on IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). It's not available with version 15.

Download local copy of scored file. If you choose to send scoring output to a flat file, you can also download a local copy of the file if desired. Administrators can use browser-based IBM SPSS Collaboration and Deployment Services Deployment Manager to enable or disable the feature, and to increase or decrease the maximum file size limit allowed. If the **Download File** button is inactive, either the file size exceeds the maximum size limit set by the administrator, or the administrator disabled the feature. The feature is disabled by default. This feature is only available with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). It isn't available with version 15.

New optional settings for building models. When building models in IBM SPSS Modeler Advantage, new optional settings are available such as data partitioning options, choosing which modeling techniques (algorithms) to include when building the model, specifying the maximum number of models to retain and combine, and setting profit criteria prior to building the model. The options available depend on the type of model and the target selected. See the topic "Optional model settings" on page 44 for more information. Note that the following new features are only available with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). They aren't available with version 15.

- Selecting which modeling techniques to include when building the model
- Setting profit criteria
- Viewing high level details about individual models
- Viewing additional details about a selected model

Tree viewer. After building a model in IBM SPSS Modeler Advantage and viewing individual model results, the name of each model is a link that opens a new dialog containing information about the model such as the target field and input fields. If the selected model is a *tree model*, a tree viewer will be displayed now also. Note that the tree viewer is only available with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). It isn't available with version 15.

Evaluation enhancements. Various enhancements have been made to the Evaluate dialog box in IBM SPSS Modeler Advantage. For example, a new **Modify Model Selection** button is available, allowing you to modify the models to include in the evaluation. After making the selections, you can also copy the current model selections back to the individual model results table, if desired. Some of the new options are only available for certain model types (most new options are only for the *Predictive* model type, and some are also for the *Predictive rule* model type). Note that the following new features in the Evaluate dialog box are only enabled with IBM SPSS Modeler Server 15 Fix Pack 1 (15.0.0.1). They aren't available with version 15.

- Performance charts for numeric target
- Profit simulation for a numeric target
- Coincidence matrix charts (for all targets)
- Table view of all charts
- The **Modify Model Selection** button
- The option to **Include individual model evaluations for performance charts**

Microsoft Windows 8 and Internet Explorer 10. IBM SPSS Collaboration and Deployment Services 5 Fix Pack 1 (5.0.0.1) adds support for Windows 8 and Internet Explorer 10. They are not supported with IBM SPSS Collaboration and Deployment Services 4.2.1 or 5.

User interface enhancements. Various minor changes have been made to the user interface to improve usability.

Version 7

IBM SPSS Decision Management 7 (7.0.0.0) added the following features:

Optimization. The goal of optimization is to identify the solution that best meets a specific goal, such as maximizing the revenue from a marketing campaign, or minimizing the risk of fraud or churn. The result of optimization is a solution that answers this question across all possible combinations within your population, in the manner that best maximizes the value you want to optimize (such as profit).

Clustering models. New in IBM SPSS Modeler Advantage, cluster modeling methods help discover groups of data records with similar values or patterns. These techniques are used in marketing (customer segmentation) and other business applications (records that fall into single-record clusters may contain errors or be instances of fraud). Clustering models are often referred to as unsupervised learning models, since they do not use a target field, and do not return a specific prediction that can be evaluated as true or false. IBM SPSS Decision Management allows you to build cluster models automatically. You can also combine, rename, modify, or delete clusters, or add custom clusters based on prior business knowledge.

Association models. New in IBM SPSS Modeler Advantage, association modeling allow you to predict which items are most likely to appear together, and predict the strength of the relationship between them. Association rules associate a particular conclusion (the purchase of a particular product, for example) with a set of conditions (the purchase of several other products, for example).

Data. The Data tab now includes the ability to have multiple data inputs, define the project data model with an Application View, refer to model outputs as if they were data, upload local data files, and preview and count data.

Improvements to IBM Cognos data integration. As with IBM SPSS Modeler 15, IBM SPSS Decision Management now supports the import of Cognos pre-built reports, Cognos filters and calculations, and control of automatic disaggregation.

Multiple elements. Applications can now include more than one dimension hierarchy. When launching an application, if the selected application is configured to include more than one element, you will be prompted to select one or more of the available elements (hierarchies) to use in the application. The dimension tree on the Define tab specifies the range of possible decisions or recommendations that can be returned. If an application includes more than one dimension hierarchy, they will be included on separate tabs of the dimension tree area on the Define tab.

Dynamic allocation. In previous releases, to allocate something, it had to exist in the dimension tree. Now, for dimension items with no children, you can read items from a table linked by a key on the project data model. For example, you might have a dimension item called *Cars* with no children. You could set up dynamic allocations on *Cars* to reference fields in a table such as *BMW*, *Audi*, and *Ford*.

Multiple allocations. A new **Multiple allocation** option allows allocation of records to multiple dimension items (such as multiple offers) for each segment. Multiple allocations can be made for each rule instead of a single allocation per rule.

User interface improvements. Many changes and additions to the IBM SPSS Decision Management user interface have been made to improve the usability and scalability of all applications.

Gallery enhancements. Various enhancements have been made to the gallery, such as a **Clear Gallery** option for removing more than one gallery item at a time.

Rule editor enhancements. Depending on where the rule editor is launched from, users can now set the rule type for selection rules (include or exclude), assign risk points for aggregation rules, select a dimension for allocation rules, and use *OR* and *Split OR* controls to combine multiple expressions within a rule and create an *OR* expression. When creating rules, now you can also select a dimension from a hierarchy tree (for applications with more than one tree) or launch the expression editor to create an expression. This functionality was not available in the rule editor dialog box previously.

Extended IBM Operational Decision Management integration. In IBM SPSS Collaboration and Deployment Services Deployment Manager, you can use the Add New External Rule dialog box to define references to IBM Operational Decision Management rulesets that can then be used in IBM SPSS Decision Management. Previously, these rules could only be accessed through a web service. With this release, you can now reference local rules in the IBM Operational Decision Management database. For the new local option, the rule is executed via the IBM SPSS Modeler Server (the same server defined in Deployment Manager for use by IBM SPSS Decision Management).

WhatIf? enhancements. Various enhancements have been made to functionality in the WhatIf? dialog box. For example, for the simulation date you can now enter range of dates on which to perform the WhatIf? evaluation.

Upload stream files. You can now open any IBM SPSS Modeler stream file from the local file system and store it to the repository if desired.

Multiple IBM SPSS Modeler Server credentials. If you have multiple IBM SPSS Modeler Server credentials defined in IBM SPSS Collaboration and Deployment Services Deployment Manager, you can now also specify more than one for use with IBM SPSS Decision Management. Then when the IBM SPSS Decision Management user logs on, the server will try the first credential in the list to see if the user has authority, proceeding through the list in order until a credential is found with the proper authority. If the user does not have authority to any of the credentials in the ordered list, he or she cannot use IBM SPSS Decision Management. This provides better control over which users have access to which parts of the IBM SPSS Modeler Server file system.

Version 6.2

IBM SPSS Decision Management 6.2 added the following features:

Support for Cognos data. When used with IBM SPSS Modeler Server 14.2 or newer, IBM SPSS Decision Management 6.2 added the ability to read data from an IBM Cognos Business Intelligence server when defining data sources, and to write scoring results back to a Cognos BI Server. After defining a connection to a Cognos BI server, you can browse packages, dimensions, queries, and reports to choose the desired fields. See the topic "IBM Cognos BI source" on page 22 for more information.

ILOG Business Rule Management System integration. Rules created in a Business Rules Management System such as ILOG can be referenced and used in IBM SPSS Decision Management applications, provided they have been developed to support the same data model used in the current IBM SPSS Decision Management project. These external rules can be selected and used in the same manner as other models and rules within the IBM SPSS Decision Management application. See the topic "Using rules from IBM Operational Decision Management" on page 36 for more information. Note that ILOG is now referred to as IBM Operational Decision Management.

Support for Linux on System Z. SUSE Enterpise Server 10 is supported on 64-bit IBM System z10.

Chapter 2. Navigating the workspace

Changing the displayed language

Each user can easily change the language of his or her application without having to restart any servers. For example, German-speaking users, English-speaking users, and Japanese-speaking users can all be using the same application in their own language.

From the **Tools** menu in your browser, go to **Internet Options** or **Options**, switch to the desired language, and then refresh the screen. If IBM Analytical Decision Management is not provided in your desired language, it will be displayed in English by default, or in your second language of choice if it is available.

The IBM Analytical Decision Management interface supports the following languages. Note that this list is subject to change from release to release, and that the documentation is not available in all the same languages as the user interface.
* English
* German
* Spanish
* French
* Italian
* Japanese
* Korean
* Portuguese (Brazilian)
* Russian
* Chinese (China/simplified)
* Chinese (Taiwan/traditional)

Launching applications

The Applications launch page provides configurable links to your installed applications. Each panel represents an installed application. Click the drop-down to choose from recently-saved projects, models, or rules. You can also add one or more custom shortcuts that point to specific, labeled versions.

* To access the Applications Launch page, navigate to *http://hostname:port/DM*, where *hostname* is the IP address or name of the repository machine and *port* is the application server port number. If you don't know this information, contact your administrator. You can also add a bookmark to this page in your browser, or create a shortcut on your desktop.

    If you have problems accessing the IBM Analytical Decision Management URL or any IBM SPSS Collaboration and Deployment Services URLs, an administrator may have customized them. Contact your administrator. More information is available in the *Application context roots* chapter of the *IBM SPSS Collaboration and Deployment Services Repository Installation and Configuration Guide*.

* To create a new project, model, or rule, select **New** from the drop-down list for the desired application and click **Go**.

* To open a saved project, model, or rule, select a recently-used version or shortcut from the drop-down list in the application panel, or select **Browse** from the drop-down list to choose a different object or version. To quickly the see the repository location of an item in the list, hover over it in the drop-down menu.

- To add a new custom shortcut that will always be available in the drop-down list, click the triangular arrow icon in the upper right corner of the application panel. Optionally, you can specify the number of shortcuts to display.

Other ways to open an application

- Copy or type the URL into the address bar in your browser, or add the URL to the list of saved bookmarks or favorites in your browser.
- Create a desktop shortcut pointing to the URL.
- Link to an application URL from within your company's Intranet site or portal.

Customizing the launch page

The controls in the upper right corner of the Application Launch page allow you to add applications, and to toggle the display between panel and list view. You can also drag panels to rearrange them on the page.

- To add a new application panel, select **Add application** in the upper right corner of the Application Launch page. (If no additional applications are available, this link is not available.)

  A list of available applications is displayed. Optionally, you can create shortcuts to one or more labeled versions.

- To change the arrangement of application panels on the Launch page, click on the title bar of any panel and hold down the mouse button to drag the panel to a new location.
- To remove an application, click the remove icon (red X) in the upper right corner of the application panel. Any shortcuts that point to that application will also be removed. (The remove icon will not be available if the application panel has been locked on the page by the administrator.)
- Unsaved projects, models, and rules may be listed in the gallery on the launch page. This gallery displays models that are still building, as well as those that have completed building but are not yet saved. See the topic "Gallery" for more information.

The Home page

The application provides a step-by-step workflow, as represented by the icons on the Home page. Click any icon to jump to that step.

Data. Define data sets for use by the application. Choose data for rule-building, simulation, and modeling.

Rules. Create, edit, and share rules for use in other applications.

Gallery

The gallery is displayed on the Applications Launch Page. It lists unsaved work, such as a model that has been left to build overnight, allowing you to start the model before you leave work, close your browser, and return to find your results in the morning. You can also use the gallery to retrieve your work after accidentally closing your browser. Items in the gallery may include projects, rules, models, data scans, evaluations, and other long-running tasks as appropriate to the application.

- If the name of the item is underlined, click the name to open the relevant item. For example, if you have an unsaved data scan, click the data scan title in the gallery to open the Data tab.
- If you do not require any item shown in the gallery, click the delete icon to remove it; a warning message is displayed to confirm you want to permanently delete the item. Or to remove several gallery items at once, click **Clear Gallery**. The Clear Gallery dialog box will be displayed allowing you to select which tabs to remove all objects from.

- On the Projects tab of the gallery, the **Created in** drop-down menu allows you to filter the items displayed on the tab for each application. For example, while working in an application you may not want to see gallery items for other applications.

Setting preferences

There are a number of preferences you can adjust to customize IBM Analytical Decision Management to your needs. Primarily, this customization consists of setting your own default options to save you having to repeatedly select the same actions when creating, testing, and saving your work. The preferences you set are unique to you and apply to any application you use.

To access the user preferences settings, from within any application, click the **Set user preferences** icon on the toolbar.

Figure 1. Set user preferences icon

You can specify your own preferred settings for each area as detailed in the sections that follow.

General settings

Default number of recently opened objects in application shortcuts. Select the number of items to be available from the drop-down list on the Application Launch page.

Date format. Select the format to use when displaying or recording dates.

Time format. Select the format to use when displaying or recording time.

Currency. Select the default currency for each new project. Note that this will not overwrite existing data brought in as a different currency.

Decimal places. Select the number of decimal places to display for numbers.

Number of rows to show on preview. Select the maximum number of records to show when displaying previews; for example, when checking a data source upon import.

Number of results per page (paginated tables). Select the maximum number of results to show on each page when displaying tables.

Maximum number of items displayed in a graph. Sometimes graph results do not work well when there are many items to display. Specify the maximum number of items to display in graphs. This will force a table to be displayed instead of a graph if the number of items exceeds this maximum. This setting only applies to categorical fields in data overviews and the scoring summary report graph for models with a categorical target.

Number of data segments (tiles) to be applied to charts and tables. Select the maximum number of data segments to use throughout the application for charts and tables, where applicable.

Test and simulation settings

Default number of records tested. Specify the default number of records to include when running a test.

Default test/simulation date. Specify whether the default simulation date is set to the current date, or a specific date.

Opening and saving projects, models, and rules

You can open from, delete from, or save to the IBM SPSS Collaboration and Deployment Services Repository. Saving to the IBM SPSS Collaboration and Deployment Services Repository provides important benefits including support for multiple versions, labeling, and security. Depending on the local configuration, not all options may be supported.

Creates a new rule.

Saves the current project to its last known location in the IBM SPSS Collaboration and Deployment Services Repository. Optionally, you can apply a new label when saving.

Saves and labels the current project, or saves under a new name. Where applicable, you can choose between saving to the IBM SPSS Collaboration and Deployment Services Repository, and downloading to the local file system.

Opens a project from the IBM SPSS Collaboration and Deployment Services Repository or the local file system as applicable. Or, where applicable, you can also upload a project from the local file system. File system access is not supported for all applications. If uploading an IBM SPSS Modeler stream, you can choose to open the file or save it to the repository..

Opening from and saving to the repository

The IBM SPSS Collaboration and Deployment Services Repository provides important benefits including support for multiple versions, labeling, and security. Each time you save you create a new version, and you can revert to an earlier version at any time. You can also manage versions using labels, and apply security settings to determine who can access, view, or remove an object.

The tree view in the left pane displays folders in the IBM SPSS Collaboration and Deployment Services Repository you have authority to see. Navigate the folder tree or perform a search to find what you're looking for. You can also filter the file list by labels available in the lower-left pane. For example, you might click a *production* label to only display files marked with that label.

To see more details about a file, when available, click the preview icon or the information icon beside it.

To delete a file, click the delete icon beside it. This is especially useful for users who do not have access to the IBM SPSS Collaboration and Deployment Services Deployment Manager client software.

Folder. Displays available folders and allows you to navigate the hierarchy. Choose the folder to save to. You can also create a new folder or rename an existing folder using the icons above the folder list.

Filter by Label. Limits the display to objects with the selected label. When browsing folders, only objects with this label are shown.

Save as. When saving rules in IBM SPSS Rules Management), a **Save as** or **Export as** drop-down is also available. Select **Single ruleset** or **Individual rules** from the drop-down to specify how to save the rule. If only referenced rules are available, the drop-down will be disabled and the rule will be saved as a single ruleset (referenced rules cannot be saved individually).

If you choose to save as individual rules, a table will be displayed to enter details for each individual rule you wish to save.

If you choose to save as a ruleset, you will be given the options to **Store with Selections** or **Store with Risk points**. If the rule you are saving is a Selection rule, you can select the **Store with Selections** option to also save whether it's an inclusion or exclusion rule. If the rule you are saving is an Aggregation rule, you can select the **Store with Risk points** option to save the aggregation rule with its aggregation points.

Name. Enter a unique name for the object you are saving.

Keywords. Enter any keywords to quickly identify the saved object when searching for it. For example these might be *Sales* and *1-2009*.

Description. Enter a longer description of the object saved. For example, this might be *Analysis of retail sales; 1st quarter of 2009*.

Label(s) If you are likely to have several versions of the same saved object, you can identify each with a unique label. Either select an existing label or create a new one.

Working with labels

Labels are used to identify a specific version of a project, model, or rule. When you share a rule between multiple projects, for example, you can specify that you want to use the version labeled *production*. If a newer version of the rule is created, the one used in your project won't change unless the *production* label is explicitly moved to that version.

When referencing shared objects, using a labeled version is strongly recommended, particularly in a production environment, as any changes to the latest version may otherwise have unintended consequences. In certain cases, such as when referencing a rule from within a rule, a labeled version must be used. See the topic "Exporting and reusing rules" on page 34 for more information.

If no version is specified, the latest version is used, which means that subsequent changes to the rule will automatically be picked up. (Think of *LATEST* as a special label that is automatically rolled to each new version as it is created.)

Each time you save a new object or version, you can choose the labels to apply to that version. See the topic "Opening and saving projects, models, and rules" on page 12 for more information.

Restrictions on label names. When defining labels for use with IBM Analytical Decision Management, the characters **? # &** and **:** (colon) cannot be included in label names, as they may prevent the application from correctly interpreting references to objects that use these labels. For example, the following are not valid label names:

Test & Deploy

DEPLOY:PHASE1

Locking projects

When you open a project, model, or rule, a Lock check box is available in the upper right corner allowing you to lock the object in the IBM SPSS Collaboration and Deployment Services Repository so that other users cannot edit it.

- When a project is locked, other users can open it for viewing, but cannot save changes.
- When you close a locked project (for example by closing the browser) the lock is released so that others may edit the object.
- If you choose to unlock a project without closing it (by deselecting the Lock check box) you will be prompted to save any changes. Unsaved changes will be lost after the object is unlocked.

- After unlocking a project, you can continue to make changes to it, unless another user happens lock it while you have it unlocked (in which case you will no longer be able to save changes).

Properties

The Properties dialog includes a **Project** tab and a **Repository** tab. To display the Properties dialog, click the Properties icon in any application.

Figure 2. Properties icon

On the **Project** tab, you can specify the maximum members allowed for *Nominal* fields. This is used on the Data tab. Select **Maximum members for nominal fields** and specify a value. The default is **250**. This is useful in situations such as the following:

1. A user creates a new data source and runs a data scan.
2. A number of fields that contain useful categorical information appear in the project data model as *Typeless* because the number of distinct values in the data exceeds the default maximum size limit set in the Properties dialog.
3. The user clicks the Properties icon and increases the **Maximum members for nominal fields** value to an appropriate size for his data.
4. Back on the Data tab, the user clicks **Refresh data scan** but the fields remain as *Typeless* because the refresh action doesn't alter the measure, to preserve any changes the user has made. In other words, a refresh retains the existing measure for a field and only recomputes the values because any changes the user has made in the data model should be preserved.
5. To get the values into the project data model, the user must manually change the measure of each field from *Typeless* to *Categorical* prior to running the refresh. Note that if there are many fields affected, it may be faster to recreate the data source.

Note that the Maximum members for nominal fields setting does not apply to data sources of the file type *.sav.*. Many *.sav* files have user customized files and values, so they will not be affected by this setting. The user must manually change the fields.

On the **Repository** tab, you can add keywords and a description to help identify a project, model, or ruleset when browsing for it at a later date:

- **Keywords.** Enter any keywords to quickly identify the saved object when searching for it. For example these might be *Sales* and *Q1-2013*.
- **Description.** Enter a longer description of the object saved. For example, this might be *Analysis of retail sales; first quarter of 2013*.

Chapter 3. Managing data sources

Use the Data tab to define data sources for analysis, simulation and testing, scoring, or other use. Data sets can be configured by your administrator, or you can add your own.

The Source tab is divided into these main areas:

- **Project Data Model.** Defines the fields required by the application. All other data sets are mapped relative to this source. The **Data Source Fields** section lists the input fields and types your project uses. If desired, use the **Derived** tab to add expressions or model output from different data sources to extend your project data model.
- **Data Sources.** The **Project Data Sources** section lists the data sources that have been saved for use with the current project, model, or rule. The **My Data Sources** section lists data sources you have defined or saved. Data sets from this list can be copied to or from different projects, and then promoted for use as the project data model, making it possible to share data across applications.

Project data model

The project data model acts like a template listing the input fields and types your project will use.

Data source. Displays the data source selected for the application.

Record ID. If desired, you can select a field to use as the *entity* ID, to enable SQL push-back. This can improve performance when the data originates in a database and if the selected field is an integer.

Operational. A check in this box shows that the data in the field is available for use in rules and expressions, and for predictions. Remove the check box if the field is to be used as the target field for analysis. For example, if you are running a query to see how well customers will respond to a direct mail campaign, the results field would be analytical since you do not know the response values in advance.

Note: If you select an IBM SPSS Collaboration and Deployment Services Enterprise View data source as the project data model, the boxes in this column are disabled.

Field name. Displays the name of each field, as listed in the data source, along with an icon that identifies its storage type.

Measurement. Displays the measurement type of each field in the data source. If you change any of the measurement types, you must refresh the data scan to ensure data compatibility. See the topic "Measurement levels" on page 24 for more information.

Values. Lists the values for each field in the data source. For example, maximum and minimum values in a range. For flag fields, hold the mouse pointer over the value to display a tool tip that indicates the "true" and "false" values defined for the field. You can also click a value to edit it.

Typically, the project data model defines the inputs as a set of fields coming from a single logical table. This is always the case for IBM SPSS Modeler Advantage. However, for other applications, the project data model may consist of a primary table along with 0 or more secondary tables. These secondary tables are defined by the secondary data sources that are associated with the primary data source.

Project data sources

Project data sources may include those predefined by the administrator, or added by users. Optionally, the administrator may have locked one or more data sources to prevent users from modifying or removing them, or locked all data options so users can't create new data sources.

If a data source's inputs don't directly match those of the project data model, you can map the former to the latter and fix the discrepancy. For example, if the project data model requires a field named *purchase* with values *Yes* and *No* (measurement level flag), then any data source used must have a comparable field that can be mapped accordingly.

Name. Displays the data source name and shows an icon that identifies its file type.

No. of records. Click the icon in this column to show the record count for any data source in the table. The count is shown beside the icon.

Preview. Click the icon in this column to preview a sample of the data contained in the source. For more information, see "Previewing data" on page 24.

Overview. Click the icon in this column for an overview of the data source. For more information, see "Data overview" on page 24.

Compatible. Either displays a note that the data source is used as the project data model, or shows a green, orange, or red ball to indicate how compatible the data source is with the project data model.
* A green ball shows that the data source is operationally compatible with the project data model data source. An operationally compatible data source is one that includes all the operational fields of the project data model, but can have additional fields. This data source is suitable for rules, scoring, simulation, and test operations.
* An orange ball shows that the data source has at least one field that is compatible with the project data model, with the same name and type. This data source may also have additional fields, and is suitable for building and evaluating models.
* A red ball shows that the data source is incompatible with the project data model, and fields must be mapped before it can be used in the application. An incompatible data source is one which has at least one field whose type is incompatible with the equivalent project data model type.

Map fields. This option allows you to compare the data source fields with those in the project data model and map or unmap any compatible fields to match those required by the project data model. For more information, see "Mapping fields" on page 25.

Copy. Copies the data source to the **My Data Sources** area.

My data sources

Data sources on this list are saved with your user account, so they are available whenever you log in, and can be copied to any project, model, or rule that you open (assuming your administrator gives you authority to do so). Fields in this part of the tab work in the same way as those in the **Project Data Sources** area; although there is no **Compatibility** column.

The **Copy** column enables you to copy the data source into the **Project Data Sources** area.

Working with data sources
* To add a new data source to either the **Project Data Sources** or **My Data Sources** lists, select **Add a data source**. For more information, see "Creating a new data source" on page 17.
* To change the measurement level, or type, of a field in the data source in the **Project Data Model**, select the relevant level. For more information, see "Measurement levels" on page 24.

- To copy data sources to or from the **Project Data Sources** list, click the appropriate arrow in the Copy column.
- To map field names for a data source to the project data model, click the appropriate link under the Compatible column. (Once field names are mapped, the link is no longer displayed.) For more information, see "Mapping fields" on page 25.
- To preview a data source, click the **Preview** icon. For more information, see "Previewing data" on page 24.

Creating a new data source

1. To access the Data Source Editor, select **Add a data source** under either the Project Data Sources or My Data Sources area.
2. Type a name for the data source.
3. Select the source type. Note that the list of available databases is from those available on the IBM SPSS Modeler machine.
   - **File.** Many file types are supported, such as text (*.txt*, *.csv*, etc.), Excel, the *.sav* and *.zsav* formats used by IBM SPSS Statistics, and SAS files. See the topic "File source" for more information.
   - **Database.** Any database that supports ODBC, including Microsoft SQL Server, DB2, Oracle, and others. See the topic "Database source" on page 20 for more information.
   - **Enterprise View.** Creates a connection to an IBM SPSS Collaboration and Deployment Services Enterprise View as defined in IBM SPSS Collaboration and Deployment Services. See the topic "Enterprise View source" on page 20 for more information.
   - **Data View.** Creates a connection to be used with IBM Operational Decision Management rules. For complex rules that use a hierarchical data model, note that administrators may restrict use of such rules because they can degrade performance. See the topic "Data View source" on page 21 for more information.
   - **IBM Cognos BI server.** Any database that supports IBM Cognos BI server. See the topic "IBM Cognos BI source" on page 22 for more information.
4. Specify options as appropriate for the source type.

   You can click the **No. of records** icon to display the number of records in the current data source, or click **Preview** to display the first ten records of the data source.

File source

The application supports several data file types. When you enter the filename or browse for the file, the application automatically detects the type and expands the Data Source Editor to display additional fields that relate to that type. If the file type is incorrect, or requires a specific version (for example, Microsoft Excel 1997-2003), you can change the type to the one required.

You can click **Browse** to select a file from the repository, or click **Upload local file** to select a file from your local file system (if enabled).

If you upload a local file, it will be uploaded to the IBM SPSS Modeler Server. In the Upload file dialog, browse to and select the local file you want to upload and then browse to and select the IBM SPSS Modeler Server upload destination. When you select the destination, you can change the file name if desired.

The **Upload local file** capability is disabled by default. Administrators can use browser-based IBM SPSS Collaboration and Deployment Services Deployment Manager to enable or disable it, and to increase or decrease the maximum file size limit allowed.

Text-based data files

When you select a text-based data source, you are prompted to enter further details.

First row has column names. Select this if the names of each column are included as a heading row in the data source.

Encoding. Specifies the text-encoding method used. You can choose between the system default or UTF-8.

Decimal symbol. Specifies how decimals should be represented in the data.
- **Period (.).** The period character will be used as the decimal separator.
- **Comma (,).** The comma character will be used as the decimal separator.

Delimiters. Using the check boxes listed for this control, you can specify which characters, such as the comma (,), define field boundaries in the file. You can also specify more than one delimiter, such as ", |" for records that use multiple delimiters. The default delimiter is the comma.

Note: If the comma is also defined as the decimal separator, the default settings here will not work. In cases where the comma is both the field delimiter and the decimal separator, select **Other** in the Delimiters list. Then manually specify a comma in the entry field.

Select **Allow multiple blank delimiters** to treat multiple adjacent blank delimiter characters as a single delimiter. For example, if one data value is followed by four spaces and then another data value, this group would be treated as two fields rather than five.

Advanced options

EOL comment characters. Specify characters, such as # or !, to indicate annotations in the data. Wherever one of these characters appears in the data file, everything up to but not including the next new-line character will be ignored.

Specify input fields. Specify the number of input fields to be used from each record.

Specify data format For **File** data source types that are of type **Variable length fields in plain text**, you can use this section to set the input storage type and the format of each field to ensure that values are read correctly. This is similar to the functionality available in IBM SPSS Modeler client. The **Override** option indicates whether the default is overridden. Selecting **Override** enables the **Storage** and **Input Format** controls. Deselecting **Override** will change the values back to their original defaults. **Input Format** only applies real, date, time, and timestamp storage types.

Skip header characters. Specify how many characters you want to ignore at the beginning of the first record.

Lines to scan for type. Specify how many lines to scan for specified data types.

Strip lead and trail spaces. Select to discard leading and trailing spaces in strings on import. You can choose to strip from the left, right, both sides, or none.

Invalid characters. Select **Discard** to remove invalid characters from the data source. Select **Replace with** to replace invalid characters with the specified symbol (one character only). Invalid characters are null characters or any character that does not exist in the encoding method specified.

Quotes. Using the drop-down lists, you can specify how single and double quotation marks are treated on import. You can choose to **Discard** all quotation marks, **Include as text** by including them in the field value, or **Pair and discard** to match pairs of quotation marks and remove them. If a quotation mark is unmatched, you will receive an error message. Both **Discard** and **Pair and discard** store the field value (without quotation marks) as a string.

Note: When using **Pair and discard**, spaces are kept. When using **Discard**, trailing spaces inside and outside quotes are removed (for example, `' " ab c" , "d ef " , " gh i " '` will result in `'ab c, d ef, gh i'`). When using **Include as text**, quotes are treated as normal characters, so leading and trailing spaces will be stripped naturally.

Specify input fields. See the topic "Selecting input fields" on page 23 for more information.

Excel data files

When you select an Excel data source, you are prompted to enter further details:

First row has column names. Select this if the names of each column are included as a heading row in the data source.

Named range. Enables you to select a named range of cells as defined in the Excel worksheet. If you use a named range, other worksheet and data range settings are no longer applicable and are disabled as a result.

Choose worksheet. Specifies the worksheet to import, either by name or by index.
- **By name.** Select the name of the worksheet you want to import.
- **By index.** Specify the index value for the worksheet you want to import, beginning with 0 for the first worksheet, 1 for the second worksheet, and so on.

Range on worksheet. You can import data beginning with the first non-blank row or with an explicit range of cells.
- **Range starts on first non-blank row.** Locates the first non-blank cell and uses this as the upper left corner of the data range.
- **Explicit range of cells.** Enables you to specify an explicit range by row and column. For example, to specify the Excel range A1:D5, you can enter A1 in the first field and D5 in the second (or alternatively, R1C1 and R5C4). All rows in the specified range are returned, including blank rows.

On blank rows. If a blank row is encountered, you can choose whether to skip and ignore the row or choose **Return blank rows** to continue reading all data to the end of the worksheet, including blank rows.

Specify input fields. See the topic "Selecting input fields" on page 23 for more information.

IBM SPSS Statistics data files

When you select an IBM SPSS Statistics data source (*.sav* or *.zsav* file), you are prompted to enter further details. If the file is password protected, you will also be prompted to enter the password.

Variable names. Select a method of handling variable names and labels upon import from a IBM SPSS Statistics *.sav* or *.zsav*file.
- **Read names and labels.** Select to read in both variable names and labels; this is the default option. Labels may be displayed in charts, model browsers, and other types of output.
- **Read labels as names.** Select to read in the descriptive variable labels from the IBM SPSS Statistics *.sav* file rather than the short field names, and use these labels as variable names.

Values. Select a method of handling values and labels upon import from a IBM SPSS Statistics *.sav* or *.zsav*file.
- **Read data and labels.** Select to read in both actual values and value labels; this is the default option.
- **Read labels as data.** Select if you want to use the value labels from the *.sav* or *.zsav* file rather than the numerical or symbolic codes used to represent the values. For example, selecting this option for data

with a gender field whose values of 1 and 2 actually represent *male* and *female*, respectively, will convert the field to a string and import *male* and *female* as the actual values.

It is important to consider missing values in your IBM SPSS Statistics data before selecting this option. For example, if a numeric field uses labels only for missing values (0 = *No Answer*, –99 = *Unknown*), then selecting the option above will import only the value labels *No Answer* and *Unknown* and will convert the field to a string. In such cases, you should import the values themselves.

Specify input fields. See the topic "Selecting input fields" on page 23 for more information.

Database source

When you select a Database source, you are prompted to enter further details:

Database. Select the database you want to use.

Connect. Select whether to connect to the database using database credentials or credentials stored in the IBM SPSS Collaboration and Deployment Services Repository. Then either Enter your database login name and password or select a credential from the repository and click **Connect** to check that you have access to the database.

Note that database login name and password are not saved. So the next time you open and run the project you will be prompted to log on, unless your administrator configured Single Sign On (SSO) for IBM SPSS Modeler Server and ODBC database connections. See the *Security Considerations* section of the IBM Analytical Decision Management installation guide for more information.

Table Name. Select the table from which data is to be used.

Strip lead and trail spaces. Select to discard leading and trailing spaces in strings on import. You can choose to strip from the left, right, both sides, or none.

Quote table and column names. Specify whether you want table and column names to be enclosed in quotation marks when queries are sent to the database (if, for example, they contain spaces or punctuation).

*   The **As needed** option will quote table and field names *only* if they include nonstandard characters. Nonstandard characters include non-ASCII characters, space characters, and any non-alphanumeric character other than a full stop (.).
*   Select **Always** if you want *all* table and field names quoted.
*   Select **Never** if you *never* want table and field names quoted.

Specify input fields. See the topic "Selecting input fields" on page 23 for more information.

IBM Analytical Decision Management supports the same database data sources as IBM SPSS Modeler for reading, writing, and pushing data. For details, see the product compatibility report for IBM SPSS Modeler 17. Note that this is different from the databases supported by IBM SPSS Collaboration and Deployment Services 7, which is the database that hosts the repository.

Enterprise View source

When you select an Enterprise View source, you are prompted to enter further details:

Application View. The drop-down list displays the selected Application View, if any. If connections have been made to other Application Views in the current session, these also appear on the drop-down list. Click **Browse** to search for other Application Views in the repository.

Table. Select the Application View table from which data is to be used.

Data provider. The drop-down list displays the names of Data Provider Definitions that reference the selected Application View. Select a Data Provider Definition or select the **-no data source-** option. Selecting the **-no data source-** option means the data source is effectively unusable in the application and will not be available in Test, Simulation, etc.

Environment. The environment value, which is set in IBM SPSS Collaboration and Deployment Services, determines which Data Provider Definitions are available.

Specify input fields. See the topic "Selecting input fields" on page 23 for more information.

For more details about the IBM SPSS Collaboration and Deployment Services Enterprise View, and complete details about how this data is configured, see the IBM SPSS Collaboration and Deployment Services documentation. Note that the IBM SPSS Collaboration and Deployment Services Enterprise View will be deprecated in a future release.

Data View source

A Data View source creates a connection to be used with IBM Operational Decision Management rules. For complex rules that use a hierarchical data model, note that administrators may restrict use of such rules because they can degrade performance. For details, see "Locking data options" on page 53.

A *Data View node* can be configured in IBM SPSS Modeler to include data in streams that was defined in an IBM SPSS Collaboration and Deployment Services analytic data view. An analytic data view defines a structure for accessing data that describes the entities used in predictive models and business rules. The view associates the data structure with physical data sources for the analysis.

For more information, see the IBM SPSS Modeler documentation and the *Source, Process, and Output Nodes* section of the IBM SPSS Modeler help.

When you select a Data View source, you are prompted to enter further details:

Analytic data source. Select the analytic data source from the repository.

Table name. Select the table from which data is to be used. This is a table from the data model in the analytic data view. Each table in the data model represents a concept or entity involved in the predictive analytic process. Fields for the tables correspond to attributes of the entities represented by the tables. For example, if you are analyzing customer orders, your data model could include a table for customers and a table for orders. The customers table might have fields for customer identifier, age, gender, marital status, and country of residence. The orders table might have fields for the order identifier, number of items in the order, total cost, and the identifier for the customer who placed the order. The customer identifier field could be used to associate the customers in the customers table with their orders in the orders table.

Data Access Plan. Select the data access plan from which data is to be used. A data access plan associates the data model tables in an analytic data view with physical data sources. An analytic data view typically contains multiple data access plans. When you change the data access plan in use, you change the data used by your project. For example, if the analytic data view contains a data access plan for training a model and a data access plan for testing a model, you can switch from training data to testing data by changing the data access plan in use.

Optional Attributes. If available and if desired, select optional attributes to add to the data source.

Specify input fields. See the topic "Selecting input fields" on page 23 for more information.

IBM Cognos BI source

You can import IBM Cognos BI database data or single list reports for use in IBM Analytical Decision Management.

Note: If the IBM SPSS Modeler Server in use with IBM Analytical Decision Management does not support IBM Cognos BI server, then this option is not available.

When you select **IBM Cognos BI server** and click **Connect** you are prompted to enter the server connection details.

1. Type the **Server URL** of the IBM Cognos server from which to import or export data. Contact your Cognos system administrator if you are not sure which URL to use.
2. Select the **Mode** in which you will connect. To log in as a specific user (for example, as an administrator), select **Set credentials** and enter the IBM Cognos namespace, user name, and password:
   - Enter the IBM Cognos security authentication provider **Namespace** used to log on to the server. The authentication provider is used to define and maintain users, groups, and roles, and to control the authentication process.
   - Enter the IBM Cognos **User name** with which to log on to the server.
   - Enter the **Password** associated with the specified user name.

   OR

   Browse to and select an IBM Cognos credential stored in the IBM SPSS Collaboration and Deployment Services Repository.
3. Alternatively, to log in with no user credentials, in which case you cannot specify the namespace, user name, and password fields, select **Use Anonymous Connection**.

   Note: Some server connections do not allow anonymous connections.

Selecting Cognos object details

After setting up the IBM Cognos server connection, select the mode. **Data** mode allows you to browse to and select a package from which to import data. **List report** mode allows you to browse to and select a predefined list report from which to import data.

A package contains an IBM Cognos model and all of the folders, queries, reports, views, shortcuts, URLs, and job definitions associated with that model. An IBM Cognos model defines business rules, data descriptions, data relationships, business dimensions and hierarchies, and other administrative tasks.

1. To select a package, select **Data** mode and click **Browse**. A dialog box is displayed listing all available packages from which you can import content.
2. Or to select a single list report, select **List report** mode and click **Browse**. A dialog box is displayed listing packages and folders from which you can select a report. *Note*: Only single list reports can be imported; multiple lists are not supported.
3. Where necessary, expand the list to find the package or report you require. Select the item and click **OK**. The package or report details are shown in the Data Source Editor.
4. After you select a package, you then choose the objects to import. For example, you select the package object query subjects (which represent database tables) or individual field query items (which represent table columns) to use from a package. *Note*: The data must be in UTF-8 format.
5. Select the **Aggregate data before importing** option if you want to import aggregated data rather than raw data. Aggregated data generally has many fewer records than raw data.

Select Objects. Displays the objects (also known as namespaces) you can select from the package or list report. Expand one or more objects to display the input fields you can import. Selecting an object automatically selects all of its input fields.

Review Parameters. If parameters have been defined, either for a data object or a report, you can specify values for these parameters before importing the object or report. This table contains the list of parameters required for filters and calculations. These may have default values. A warning is provided if one or more parameters does not have a default. Set parameters as appropriate in the **Value** column. You can use parameters to make adjustments (for example, to perform a parameterized calculation) before importing the data.

Review Filters. If a package has filters defined, you can import one or more of them. If a filter you import is associated with imported data, that filter is applied before the data is imported. Filters may or may not be parameterized.

Review Input Fields. Lists the input fields (also known as database objects) that will be imported into IBM Analytical Decision Management for processing. If you do not require a particular field, deselect the check box next to its field name.

Cognos object icons: The various types of objects you can import from a Cognos BI database are represented by different icons, as the following table illustrates.

Table 1. Cognos object icons.

Icon	Object
	Package
	Namespace
	Query subject
	Query item
	Measure dimension
	Measure
	Dimension
	Level hierarchy
	Level
	Filter
	Report
	Standalone calculation

Selecting input fields

When you have a large amount of data, you may want to reduce or fine-tune the data imported from your source (for example, to only use the fields that are relevant to your area of interest).

1. At the bottom of the Data Source Editor dialog, click **Specify Input Fields**; the dialog expands to display all of the data fields in your source. In addition, all expressions you create are also displayed. For details about creating expressions, see "Expression Builder" on page 26.

2. Select each of the fields and expressions you want to use.

In addition to selecting data you can, where required, change the measurement level for one or more of the data fields. For more information, see "Measurement levels."

Previewing data

To verify that you selected the correct data source, click the **Preview** icon to display a sample of the data that is contained in the source.

Data overview

The data overview feature is available on the Data tab of all applications, on the Modeling tab and the Score tab in IBM SPSS Modeler Advantage, and on the Deploy tab of applications that include the Score Now feature.

1. To run an overview of a data source, click the Data Overview icon available throughout the application.

Figure 3. Data Overview icon

2. The Data Overview dialog will appear. If desired, select an overlay field from the drop-down in the Overview Options section.

   Then after running the overview, tabs will be available to display results for the primary selected field only, or to overlay the primary selected field with the field specified in this drop-down. For example, in the results you may want to view information about the **Age** of customers, and then overlay it with another field such as **Gender**.

3. Select the data fields to include in the overview and click **Run Overview**. All fields available in the data source are listed. All fields are selected by default.

   The data overview will run and the results will appear. You can sort the information and choose which columns to display.

4. Click any field to see its details. A new results tab will open for each field you select, allowing you to view charts and tables and select overlay fields, if available.

Measurement levels

The measurement level helps characterize the information represented by each data field, and may determine how a given field is used in rules, modeling, or other applications. The measurement level can be specified in the Project Data Model for a data source; for example, you may want to set the measurement level for an integer field with values of 0 and 1 to *Flag*, to indicate that 0 = *False* and 1 = *True*. Alternatively, you can change the level in the **Data Source Editor** dialog when you specify the input fields to be used. For more information, see "Selecting input fields" on page 23.

The following measurement levels are available:

* **Continuous** Used to describe numeric values, such as a range of 0–100 or 0.75–1.25. A continuous value can be an integer, real number, or date/time.
* **Categorical** Used for string values when an exact number of distinct values is unknown. This is an **uninstantiated** data type, meaning that all possible information about the storage and usage of the data is not yet known. Once data have been read, the measurement level will be *Flag*, *Nominal*, or *Typeless*, depending on the maximum number of members for nominal fields specified in the Project Properties dialog box.

- **Flag** Used for data with two distinct values that indicate the presence or absence of a trait, such as `true` and `false`, `Yes` and `No` or `0` and `1`. In IBM Analytical Decision Management, the first value from the category list will be used as "false" value and the second will be the "true" value, as indicated by a tool tip when hovering over the values when displayed on the Data tab. Data may be represented as text, integer, real number, date, time, or timestamp.
- **Nominal** Used to describe data with multiple distinct values, each treated as a member of a set, such as `small`/`medium`/`large`. Nominal data can have any storage—numeric, string, or date/time. Note that setting the measurement level to *Nominal* does not automatically change the values to string storage. For information about setting the maximum members allowed for nominal fields, see "Properties" on page 14.
- **Ordinal** Used to describe data with multiple distinct values that have an inherent order. For example, salary categories or satisfaction rankings can be typed as ordinal data. The order is defined by the natural sort order of the data elements. For example, 1, 3, 5 is the default sort order for a set of integers, while `HIGH`, `LOW`, `NORMAL` (ascending alphabetically) is the order for a set of strings. The ordinal measurement level enables you to define a set of categorical data as ordinal data for the purposes of visualization, model building, and export to other applications (such as IBM SPSS Statistics) that recognize ordinal data as a distinct type. You can use an ordinal field anywhere that a nominal field can be used. Additionally, fields of any storage type (real, integer, string, date, time, and so on) can be defined as ordinal.
- **Typeless** Used for data that does not conform to any of the above types, for fields with a single value, or for nominal data where the set has more members than the defined maximum. It is also useful for cases in which the measurement level would otherwise be a set with many members (such as an account number). When you select **Typeless** for a field, the role is automatically set to **None**, with **Record ID** as the only alternative. The default maximum size for sets is 250 unique values. This number can be adjusted or disabled in the Project Properties dialog box, which can be accessed from the toolbar icon.
- **Collection** Used to identify non-geospatial data that is recorded in a list. A collection is effectively a list field of zero depth, where the elements in that list have one of the other measurement levels.
- **Geospatial** Used with the List storage type to identify geospatial data. Lists can be either List of Integer or List of Real fields with a list depth that is between zero and two, inclusive.

Editing data field values

In data tables, the **Values** column lists the values for each field in the data source (for example, maximum and minimum values in a range). For flag fields, hold the mouse pointer over the value to display a tool tip that indicates the "true" and "false" values that are defined for the field. You can also click a value to edit it. A dialog box opens, and its contents vary depending on the type of data field selected. For example:
- If the field type is **Continuous**, you can enter upper and lower boundary values.
- If the field type is **Flag**, you can swap the true and false values.
- If the field type is **Nominal**, you can add, remove, reorder, and rename values.

Mapping fields

When adding a new data source, you may need to map all of its fields to ensure compatibility with the project data model. For example, if the project data model requires a field named *purchase* with values *Yes* and *No* (measurement level flag), then any data source used must have a compatible field. If the field names are not identical, they can be mapped accordingly. Note that the input and associated mapped field must have the same data type.

Alternatively, if the project data model changes, then it may be that many data sources no longer map to it correctly. If that is the case, then the compatibility indicator ball in the **Project Data Sources** area will change to the appropriate color for each of those data sources; you will then have to remap all of them to the new project data model.

1. In the **Project Data Sources** area, click **Map fields**. A dialog opens, and the system makes a best guess at mapping fields for you. At any point, you can return to the system's best guess by clicking **Reset**.
2. Select a field in the **Available Inputs** column, and its corresponding field in the **Required Input Fields** column, and click **Map** to map the two fields. Repeat this until all fields in the **Available Inputs** column have been mapped.
3. When finished, click **OK**.

When possible, IBM Analytical Decision Management will automatically map compatible storage types when a user adds an additional data source. If the type is not compatible, the user will be allowed to convert types.

The following table lists the allowed mappings.

Table 2. Types that can be converted

To →						
From ↓	**Int**	**Real**	**String**	**Date**	**Time**	**timestamp**
Int	n/a	Auto	Yes	Not allowed	Not allowed	Not allowed
real	Yes (decimal digits truncated)	n/a	Yes	Not allowed	Not allowed	Not allowed
String	Yes	Yes	n/a	Yes	Yes	Yes
Date	Not allowed	Not allowed	Yes	n/a	Not allowed	Yes (time as 00:00)
Time	Not allowed	Not allowed	Yes	Not allowed	n/a	Yes (base date)
Timestamp	Not allowed	Not allowed	Yes	Auto (time truncated)	Auto (date truncated)	n/a

Expression Builder

You can use the expression builder to derive additional fields or attributes for use by the application. For example, if you are using banking data, you may want to create an expression that shows the ratio between a customer's income and the number of loan accounts they have.

Expressions are available throughout the application in the same way as operational data (for example, in rules and in model predictors).

Creating an expression
1. In the Expression Builder dialog, enter a unique name for the expression.
2. Type in the **Expression Equation** field, using the function and field lists as references

   or
3. Select the desired fields and functions from the lists.
4. Double-click or click the blue arrow to add the field or function to the **Expression Equation** field.
5. Use the operand buttons in the center of the dialog to insert operations into the expression.
6. Enter a **Description** that explains what the expression is used for.
7. Click **Check** to validate the expression, then click **Save**.

Note: You can copy and paste expressions from IBM SPSS Modeler client, but note that the results may not always be as expected if the expression is stream-dependent (for example, if you paste a supported @ function or row function, it may not produce the desired results because the rest of the stream may not be set up to support it).

The only expressions that are officially supported are the ones provided in the IBM Analytical Decision Management user interface.

Selecting functions

The function list displays all available functions and operators. Scroll to select a function from the list, or, for easier searching, use the drop-down list to display a subset of functions or operators. Available functions are grouped into categories for easier searching.

Following are details about the categories. For more detailed information, refer to the IBM SPSS Modeler documentation.

Table 3. Function categories

Function Type	Description
Recently Used Functions	Contains a list of functions used during the current session.
Operators	Lists all the operators you can use when building expressions. Operators are also available from the buttons in the center of the dialog
Information Functions	Used to gain insight into field values. For example, the function `is_string` returns true for all records whose type is a string.
Conversion Functions	Used to construct new fields or convert storage type. For example, the function `to_timestamp` converts the selected field to a timestamp.
Comparison Functions	Used to compare field values to each other or to a specified string. For example, `<=` is used to compare whether the values of two fields are lesser or equal.
Logical Functions	Used to perform logical operations, such as `if`, `then`, `else` operations.
Numeric Functions	Used to perform numeric calculations, such as the natural log of field values.
Random Functions	Used to randomly select items or generate numbers.
String Functions	Used to perform various operations on strings, such as `stripchar`, which allows you to remove a specified character.
Datetime Functions	Used to perform various operations on date, time, and timestamp fields.
Blanks And Null Functions	Used to access, flag, and frequently fill user-specified blanks or system-missing values. @BLANK is not supported (because its behavior is dependent on other nodes like the Type node). `@NULL(FIELD)` and `undef` are supported. `@NULL(FIELD)` is type Boolean and returns true if the value of FIELD is the system-missing `$null$`. Returns false for all other values, including user-defined blanks. If you want to check for both, use `@BLANK(FIELD)` and `@NULL(FIELD)`. `undef` is type Any and is used generally to enter a `$null$` value—for example, to fill blank values with nulls in the Filler node.
All Functions	Contains a complete list of available functions.

Selecting fields and values

The field list displays all fields available. Scroll to select a field from the list. Double-click or click the blue arrow to add a field to the expression.

Viewing and selecting values

Field values can be viewed from a number of places in the application. To view values for a field from the Expression Builder, select the desired field and click the expression inputs button to open a dialog

box listing values for the selected field. You can then select a value and click **Insert** to paste the value into the current expression.

Figure 4. Expression inputs button

For flag and nominal fields, all defined values are listed. For continuous (numeric range) fields, the minimum and maximum values are displayed.

Checking expressions

Click **Check** in the Expression Builder (lower right corner) to validate the expression. Expressions that have not been checked are displayed in red. If errors are found, a message indicating the cause is displayed.

The following items are checked:
- Correct quoting of values and field names
- Correct usage of parameters and global variables
- Valid usage of operators
- Existence of referenced fields
- Existence and definition of referenced globals

If you encounter errors in syntax, try creating the expression using the lists and operator buttons rather than typing the expression manually. This method automatically adds the proper quotes for fields and values.

Note the following limitations when building expressions in IBM Analytical Decision Management. Expressions cannot contain any of the following items:
- A reference to an IBM SPSS Modeler stream parameter
- A reference to an IBM SPSS Modeler stream global
- A reference to a database function
- A reference to one of the following special field or field value @ functions:
  - @TARGET
  - @PREDICTED
  - @FIELD
  - @PARTITION_FIELD
  - @TRAINING_PARTITION
  - @TESTING_PARTITION
  - @VALIDATION_PARTITION

Using expressions that return a Boolean

IBM SPSS Modeler and IBM Analytical Decision Management do not have a Boolean data type. Boolean expressions cannot be used to directly produce a new field. The use of Boolean expressions is limited to expression constructs that expect a boolean expression (such as the `if` statement). Following are three examples:

```
if ('RFM Score' > 5) then "T" else "F" endif
if ('RFM Score' < 5) then "Low" elseif  ('RFM Score'  >= 5 and 'RFM Score' < 10) then "Medium" else "High" endif
hassubstring('Phone type','Coolphone')
```

In the first two examples above, the "RFM Score" expressions (`'RFM Score' > 5`, `'RFM Score' >= 5`, and `'RFM Score' < 10`) can be thought of as boolean.

These expressions cannot be used by themselves to derive a field. An error such as `Invalid type for an expressions which is not a condition: Boolean` would be displayed. But expressions like these can be used in conditional expressions, as in the full examples, and conditional expressions can be sued to derive fields with categorical measurements such as a Flag type field. Flag is the default measurement for a field with two values.

The second example above could generate a field with three distinct values (as long as records which fall into each of the Low, Medium, and High categories are found during the data scan). This would likely result in the derived field's measurement level defaulting to Nominal. The third example above is unlikely to return a field with two distinct values.

See "Measurement levels" on page 24 for more information about measurement levels.

Chapter 4. Working with rules

Rules are used to automate decisions based on business logic, or on the output from predictive models. Depending on the application, you might use rules to exclude customers based on payment history, to refer risky claims to a special investigations unit, or to target those with the highest propensity to churn, buy, or commit fraud as identified by a predictive model. While rules bring automation, models provide predictive power, and the ability to combine them is one of the key features that distinguish predictive applications from other analytical tools.

Rules can be created, applied, and saved at different points within each application. Alternatively, you can create and edit shared rules using IBM SPSS Rules Management.

IBM SPSS Rules Management

IBM SPSS Rules Management provides a central tool for creating and editing shared rules, which can be used throughout applications to select and process records, and to automate decisions accordingly. While rules can also be created and saved within each application, IBM SPSS Rules Management allows you to save rules as separate objects that can be referenced from multiple applications. For example, a rule to exclude customers who are below legal age could be shared across all applications to enforce a global policy, and any changes to the shared rule would be applied across all applications.

Note that IBM SPSS Rules Management may not be installed at all sites or available to all users. As an alternative, rules can also be defined and saved locally for use within each application.

Creating rules

- In IBM SPSS Rules Management, define or choose a data source, and select it for the project data model. See the topic Chapter 3, "Managing data sources," on page 15 for more information.

  If you open an existing rule, the associated data source will automatically be added to the data list. If changing the data source for an existing rule, the new source must be compatible with the original source to ensure the required inputs are available.

- In IBM SPSS Rules Management, specify the type of rule as either segment, selection, or aggregation.
- At the point in the application where you want to add a rule, click the **Create a new rule** icon.
- Alternatively, click the **Find an existing rule** icon to browse for existing rules which can be added or reused in the current application. See the topic "Exporting and reusing rules" on page 34 for more information.
- To create a shared rule, launch IBM SPSS Rules Management from the Application Launch page.

Defining segment rules

Segment rules are used throughout the application to select and allocate records in different applications as appropriate. Segments are defined by one or more expressions that return a true or false result. For example, you might have a rule with a single segment that returns true if the value of the age field is less than 18. Segments can be defined based on fields or expressions defined in the data, or on the propensity scores or predictions returned by a model. In applications with a Define tab that are configured to have multiple dimension hierarchies (referred to as *elements*), you can also choose from other dimension members.

Each segment is defined by one or more expressions, and returns a value of true if all conditions are met. For example, a segment defined by age and gender would only include records that met both conditions, and the rule would return *false* for all others.

To create or edit a segment rule

1. In the application where you want to add the rule, click the **Create a new rule** icon.
2. Specify a name for the rule.
3. For the input column, click the down-arrow icon to specify a field, rule, or model, as desired.

Table 4. Input column options.

Icon	Usage
	If enabled, allows you to select a dimension member from a previously defined element.
	Allows you to choose a field from the current data set.
	Allows you to choose an existing model or rule from the repository.
	Allows you to build a new model.
	Allows you to build an expression.

4. Choose the function you want to use. See the following table for available options.

Table 5. Available operators.

Value type	Available operators
Time stamp, Date, Date time	• = (equals) • != (not equal to) • > (greater than) • < (less than) • >= (greater than or equal to) • <= (less than or equal to) • BETWEEN • ISNIL • IS NOT NIL
A range	• = (equals) • != (not equal to) • > (greater than) • < (less than) • >= (greater than or equal to) • <= (less than or equal to) • BETWEEN • ISNIL • IS NOT NIL • ONE OF • NOT ONE OF

Table 5. Available operators (continued).

Value type	Available operators
A set	= (equals) != (not equal to) - ISNIL - IS NOT NIL - ONE OF - NOT ONE OF
A flag	= (equals) != (not equal to) - ISNIL - IS NOT NIL

5. Specify a value as appropriate for the input type. For non-numeric values, a dropdown list of choices is provided.

Alternatively, click the **Find an existing rule** icon to browse for rules that have already been defined. See the topic "Exporting and reusing rules" on page 34 for more information.

Note that a segment rule such as `Age BETWEEN 50 and 20` will not return results. It is the same as the expression `x <= value and value < y`, which can never be true. This is consistent with SQL Server behavior. `Age BETWEEN 20 and 50` is the correct format.

Also note that a segment rule such as `Gender = $null$` will not return results either. If you want to filter out the fields for which the `Gender` value is empty, you can instead use an expression like `Gender Not One Of (Male, Female)`.

Segment rule sets

- Segment rules can also be combined into sets. A segment ruleset returns a value of true if *any* of the individual segments is true. Note this differs from the expressions *within* a segment, all of which must be true.
- You can export segment rulesets from any application that uses rules, or you can do this in IBM SPSS Rules Management if available. See the topic "Exporting and reusing rules" on page 34 for more information.

Defining selection rules

A selection rule is a set of one or more segment rules or rule sets with an *Include* or *Exclude* attribute assigned to each. Selection rules are used throughout the applications to select subsets of records for modeling, or scoring, or other processing.

1. In the application where you want to add the rule, click the **Create a new rule** icon. This icon is available at points where selection rules can be applied.
2. Specify a name for the rule, and add one or more expressions as desired. See the topic "Defining segment rules" on page 31 for more information.
3. Click **OK** to save the rule.

Alternatively, click the **Find an existing rule** icon to browse for rules that have already been defined. See the topic "Exporting and reusing rules" on page 34 for more information.

Evaluating selection rules

Click the **View selection rule logic** icon to see details about how selections are made (logical AND versus OR).

- When evaluating selection rules, if multiple include segments are specified, records matching *any* of the segments are included. For example, customers might be included based on credit rating OR tenure.
- If multiple exclude segments are specified, records matching any of the segments are excluded. For example, customers might be excluded based on age OR credit rating.
- In case of a conflict, exclude rules take precedence. For example, a customer that falls into any excluded segment will be excluded regardless of any segments they would otherwise qualify for.
- When creating selection rule sets in IBM SPSS Rules Management, all rules must be of the same type—either *Include* or *Exclude.* You cannot combine inclusion and exclusion rules in the same rule set.

Defining aggregation rules

Aggregation rules allow you to sum one or more measures across a set of segment rules. For example, you can assign risk points based on the number of risk factors that apply. The sum of each measure for all true segments is included in the result.

1. In the application where you want to add the rule, click the **Create a new rule** icon available on the Define tab in applications where aggregation rules can be applied.
2. Specify a name for the rule, assign it risk points (or do this later), and add one or more expressions as desired. See the topic "Defining segment rules" on page 31 for more information.
3. Click **OK** to save the rule, and repeat as needed to add additional segments.
4. In the aggregation rule editor, specify the number of risk points assigned to each segment if you have not done this already.
5. Specify the number of risk points assigned to the remainder. Using the drop-down list toward the upper right, you can also specify whether this value is added to all records, or only in cases where no other rules apply.
6. Optionally, you can export aggregation rulesets for use in other applications. See the topic "Exporting and reusing rules" for more information.
7. Optionally, click the **Add annotations** icon to add a column into which you can enter text to be returned with the rule. See the topic "Adding annotations" on page 36 for more information.
8. Optionally, use the arrows to change the order in which items are displayed. The order has no effect on rule execution as the aggregation returns the same value regardless of order.

Combine/split rules

You can select one or more rules and click **OR** to combine multiple rules into a single OR statement that assigns a specified number of points if *any* of the conditions is met.

To split an OR statement so that each condition is evaluated separately, select the statement and click **Split OR**.

Exporting and reusing rules

By default, rules created in an application are saved with that application. Alternatively, rules can be exported or saved to IBM SPSS Collaboration and Deployment Services Repository for use in multiple applications. In this case each application stores a reference to a specific, labeled version of the shared rule; any changes to the rule are automatically applied to all applications that use that version.

Exporting rules

From any application that contains a local rule you want to share, click the **Export** icon to save the rule to IBM SPSS Collaboration and Deployment Services Repository where it can be reused as needed.

- When exporting multiple rules, specify whether you want to export them all in a single ruleset, or individual rules. In the case of individual rules, each segment is exported as a separate rule.
- When exporting multiple selection rules, if you choose **Store with Selections**, a selection ruleset is exported. In this case you can export either inclusions or exclusions, but not both. If you need a mix of inclusions and exclusions, these need to be exported in separate rulesets. If **Store with Selections** is not selected, a segment ruleset is exported.
- When exporting from an aggregation rule editor, if you choose **Export with risk points**, an aggregation ruleset is created. If this option is not selected, a segment ruleset is exported.

Using existing rules

- In any application where you want to make use of a shared rule, click the **Find an existing rule** icon to browse IBM SPSS Collaboration and Deployment Services Repository for existing rules that can be reused in the current application.
- To reference a specific labeled version, click the arrow next to the rule name and choose the desired version. If no version is selected, the latest is used. (Think of *LATEST* as a special label that is automatically rolled to each new version as it is created.) Using a specific labeled version is recommended to avoid unexpected results that can occur when a new version of a rule is created.

  Referenced rules can be combined with local rules in many cases, but are visually distinguished from local rules by bold text, and by an icon that indicates the type of rule (or ruleset). Referenced rules cannot be edited directly in the application where they are used, but can be edited in IBM SPSS Rules Management, which can be access from the Application Launch page. Note IBM SPSS Rules Management may not be installed at all sites or available to all users.

- Alternatively, referenced rules can be converted to local rules and saved in the current application if desired. To do this, click the name of any referenced rule, and then select Convert to single rules in the Rule Preview dialog box.

Guidelines for sharing rules

When browsing and working with shared rules, the following guidelines apply:

- Rules can be exported for reuse from any application, but once exported, can be edited in IBM SPSS Rules Management, and cannot be edited in any other application. Note that IBM SPSS Rules Management may not be installed at all sites or available to all users.
- Using a specific labeled version (other than *LATEST*) is strongly recommended, particularly in a production environment, as ongoing changes the latest version may otherwise have unintended consequences. In certain cases, such as when referencing a rule from within a rule, a labeled version must be used.
- Single rules can be added from IBM SPSS Collaboration and Deployment Services Repository by reference, or converted to local rules from the Rule Preview dialog box.
- Single rules can be referenced and combined with local rules in any rule editor. If exported as a rule set, nested references to single rules will be preserved.
- Segment and selection rule sets can be referenced and combined like single rules in most cases, but the combined rule set cannot be exported without first converting any nested rule sets to single rules. Thus an exported rule set can contain references to nested rules, *but not to nested rule sets*.
- Aggregation rule sets can also be referenced, but cannot be combined with other rules, and cannot contain nested references to other rules. When referencing an aggregation rule set, the referenced rule replaces the entire contents of the aggregation rule editor. (Alternatively, if not referencing an aggregation rule set, the aggregation editor can contain a mix of local rules, referenced segment rules, and segment rule sets.)

- When using IBM SPSS Rules Management, the same restrictions apply as for exported rules. Nested rules are allowed, nested rule sets are not. This is because IBM SPSS Rules Management is designed to create and edit shared rules (similar to exporting from any of the other applications).
- To avoid any of the restrictions on referenced rule sets, whether in IBM SPSS Rules Management or elsewhere, a rule set can always be converted to separate, local rules.
- Shared rules cannot contain references to external IBM Operational Decision Management rules.

Adding annotations

In an aggregation or allocation, click the **Add Annotations** icon to add a column of free text fields into which you can enter a note that will be included with the rule output. For example, this may be extra information that tells a call center operator who to contact when an insurance claim is flagged for referral.

Annotations can be used in combination with the **Return with** value specified for a dimension, to provide additional detail. While the **Return with** value is associated with the dimension or decision that is returned, the annotation tells you which rules fired as input to that decision.

Using rules from IBM Operational Decision Management

Rules created in a Business Rule Management System such as IBM Operational Decision Management be referenced and used in IBM Analytical Decision Management applications. These external rules can be selected and used in the same manner as other rules by clicking the **Find an existing rule** icon at the desired point in the application. External rules are distinguished by their icon, but otherwise usable in the standard manner.

For information on creating external rules for use in IBM Analytical Decision Management, see the *Application Designers Guide*.

The rule is inserted into the application in the same manner as any other rule. Note that rule outputs must match the context in which the rule is used; for example allocations require a true/false or yes/no (boolean) value as input to the decision.

As with other rules, click the name of any external rule used in the application to view details on the rule including input and fields, along with the rule description and other properties.

If an external rule has multiple output fields, a prompt is displayed allowing you to choose the desired field.

If an external rule is used in place of a model, no target field is shown, and all configured outputs are shown in the drop-down list.

Tips for using external rules
- Rules that return a value of *true* or *false* (or *yes/no* or *0/1*) can be used in any selection or allocation, as input to any rule in the rule editor, or as a **Return with** field associated with a specific dimension item.
- Rules that return categorical outputs (*red*, *white*, or *blue*; *north*, *south*, *east*, *west*, etc.) can be used in the **Use a model to decide** section on the Define tab, as an input in the rule editor, or as a **Return with** field associated with a dimension item.
- Rules that return numerical outputs can be used in the **Use a model to decide** section on the Define tab, as input to a prioritization, as an input in the rule editor, or as a **Return with** field associated with a dimension item.

- Analytic data view data sources support the use of complex (hierarchical) external rules. To reference a complex external rule, the project data model must be a *Data View* source and the functionality must be enabled (on the Data tab, an administrator must select the appropriate **Allow use of ODM rules based on hierarchical data** option).
- You cannot create a shared rule in IBM Analytical Decision Management that contains a reference to an external rule.

Testing the rule

The Test dialog box allows you to see how one or more sample records will be scored. Results are specific to the context in which the test is run, and may reflect the scores returned by a model or rule. Following are some examples of the types of results you may see, depending on your application type:

- You might look at a handful of specific customers and see which offers are allocated to each of them.
- You might look at a handful of customer records and see which customers are likely to respond to a direct mail campaign.
- You might look at a handful of specific insurance claims and see which actions are allocated to each of them.
- You might look at a handful of specific cases and see how each rule evaluates.

1. On the Define tab, click **Test**. The Test dialog is displayed.
2. Select the data source you want to use, or select **Specify custom data** in the Records section to enter test records manually.
3. Specify selections to use, if any. For example, you may want to include or exclude certain records from the test. You can search for existing rules, or create new ones as appropriate. See the topic "Defining selection rules" on page 33 for more information.
4. If using a data source, specify the number of records you want to return. You can return the first *N* records, up to a maximum of 100, read from the top of the data or from a random sample of records.
5. If using custom data, select **Specify custom data** and click **Add custom record**. All fields defined in the project data model are listed, allowing you to specify a value for each. Click **Save** when finished.
6. Under **Select Fields to Display**, select the fields you want to appear in the test output.
7. Click **Run**. The test summary displays a list of successful allocations for each record.
8. Click the information icon at the beginning of each row to obtain details for that row.

Chapter 5. Building predictive models

About predictive models

Models can be used to predict what is likely to happen in the future, based on patterns in past data. For example, you might use a model to predict which customers are least likely to churn, or most likely to respond to a particular offer, based on characteristics such as income, age, and the organizations and memberships they subscribe to.

Models can be used in the same way as rules, but while rules may be based on corporate policies, business logic, or other assumptions, models are built on actual observations of past results, and can discover patterns that may not otherwise be apparent. While rules bring common business logic to applications, models lend insight and predictive power.

Data for building and scoring models

Two different types of data are used in the modeling process:

- To build the model, you need information about the thing you want to predict. For example, if you want to predict churn, you need information about customers who have churned in the past. This is often referred to as historical or analytical data, and must contain some or all of the fields in the project data model, plus an additional field that records the outcome or result you want to predict. This extra field is used as the **target** for modeling.

- To use the model in predicting future results, you need data about the group or population you are interested in, such as potential customers or incoming claims, for example. This is often referred to as **operational** data, or **scoring** data. The project data model is typically based on this data.

- If the target field is included in the data model, the **Operational** column should not be selected for this field, because it is not available when scoring the model. Having built the model using historical data, the goal in scoring is to apply the model to new data where the outcome is not already known.

Building models

You can build a new model using any of the following methods:

- When creating or editing a rule, click the **Build a new model** icon from the input field toolbar in the rule editor.

Figure 5. Build model icon

Building a predictive model

When building a predictive model, the only required specifications are a data source and target field. You can also choose options for preparing and partitioning data, and for selecting the fields and records to use. The model is generated by applying a range of methods as appropriate to the target type, and automatically identifying the specific technique or combination that performs best on the selected data.

When scored, the model returns one or more fields containing predictions, as well as confidence values associated with those predictions. For example, if the target field is a flag field named *Response*, the model returns a field named *XF-Response* containing predicted values for that field.

1. Specify a data source. This can be any analytical data source that records the outcome you want to predict.

2. Specify a target field that records the result you want to predict. For example, the target field might indicate which customers have churned, responded to past offers, raised fraudulent claims, and so on.

   All other fields will automatically be included as inputs that may be useful in predicting this value.

3. Specify optional settings as desired. See the topic "Optional model settings" on page 44 for more information.

4. If desired, click the Data Overview icon to see an overview of the data that will be used to build the current model. See the topic "Data overview" on page 24 for more information.

5. Click **Build Model**.

   You can close the browser or work on other tasks while the model builds. The model can be accessed from the Gallery once building is complete. See the topic "Gallery" on page 10 for more information. If the model is taking longer to build than expected, click the **Stop** button to revert to the previously-saved version.

   When model building is complete, results are displayed. You can choose to view the **Combined model results** (performance charts for the combined model that was currently built) or the **Individual model results** (high level details of the individual models that make up the combined model).

   When viewing *individual model results*, the name of each model is a link that opens a new dialog containing information about the model such as the target field and input fields. If the selected model is a *tree model*, then a tree viewer will also be displayed.

6. Optionally, use the Evaluate and Test features to see how the model performs on your sample data. See the topic "Evaluating models" on page 47 for more information.

7. Save the model before closing the model builder or returning to the application.

8. Click **Use Model**, and select the model field you want to use. For example if you want to use the value predicted by the model as input to a rule, select the field that contains the predictions.

Predictive modeling results

When a predictive model completes, the results section allows you to switch between viewing **Combined model results** and **Individual model results**. The combined model results include distribution and variable importance charts for the combined model that was built, providing a quick visual assessment of model performance and of the importance of different predictors in obtaining the final result. The individual model results list information about all the individual models that make up the combined model, allowing you to click each individual model to view its details.

Distribution charts

The model distribution charts display observed values on the horizontal axis, with predicted values on the vertical axis or overlay. This allows you to see how often each value or response was predicted, and how often that prediction was correct, providing a quick visual assessment of model accuracy.

- For flag or set targets, the distribution is shown using a bar chart, with one bar for each category of the target field. The height of each bar indicates how many times that result is predicted by the model. The colored regions within each bar indicate how many times the matching result was actually observed in the testing data. If the model is reasonably accurate, then the largest area within each bar should match the corresponding response.

- For range targets, the distribution is represented by a binned scatterplot. Again, predicted values are plotted on the horizontal axis, and observed values are plotted on the vertical axis. Ideally the dots should cluster along a diagonal line. The more scattered the dots (results) and the farther from a linear distribution, the less accurate the model is.

Predictor importance charts

Typically, you will want to focus your modeling efforts on the predictor fields that matter most and consider dropping or ignoring those that matter least. The predictor importance chart helps you do this by indicating the relative importance of each predictor in estimating the model. Since the values are

relative, the sum of the values for all predictors on the display is 1.0. Predictor importance does not relate to model accuracy. It just relates to the importance of each predictor in making a prediction, not whether or not the prediction is accurate.

Calculating predictor importance may take significantly longer than model building, particularly when using large datasets.

If the option to automatically partition data was selected when the model was built, then the testing partition will automatically be selected on the Evaluate page.

If partitioning was not turned on, you will need to select a data source for evaluating the model. Selecting a data source other than the one used to build the model is recommended. Note that partitioning is on by default, so this should not be necessary unless you have changed this setting.

Individual model results

Individual model results show information about each individual model that makes up the combined model. The name of each model is a link that opens a new dialog containing full details about that model, such as the target field and input fields. If the selected model is a *tree model*, then a special tree viewer will also be displayed. The tree viewer is interactive and provides the following functionality:

- Zoom in and out on the tree
- Show the tree items as tables, graphs, or tables and graphs combined
- Change the tree's orientation to top-down, left-right, or right-left
- Use the *tree map window* and the arrow icons to navigate large trees more easily

Building a predictive rule model

Predictive rule models allow you to identify segments that show a higher or lower likelihood relative to a specific outcome. For example, you might look for customers who are least likely to churn, or most likely to say yes to a particular offer. The resulting model resembles a list of selection rules, with an **Include** or **Exclude** attribute that specifies how each segment is scored.

Predictive rule models resemble selection rules, but differ in how they are scored. In the case of a selection rule, excluded records are removed, resulting in fewer records in the output. In the case of a model, no records are removed, but additional fields are added indicating the prediction and propensity values for each record. Included segments are scored as *True* (or *False*, or whatever outcome you are trying to predict). Excluded segments are scored as *NULL*, as are any records that fall to the remainder because they do not match any rule.

Each record is also given a propensity score that indicates the likelihood of the target response, making it possible to identify the customers who are most likely to respond. As such, predictive rule models are well-suited for generating mailing lists, and they are widely used in customer relationship management, including call center or marketing applications.

- Predictive rule modeling is available for categorical targets only.
- You can view, modify, and reorganize segments and to choose which to include or exclude for purposes of scoring. For example, you can choose to exclude one group of customers from future offers and include others and immediately see how this affects your overall hit rate.
- As with selection rules, you can manually add or edit segments as needed. It is often easier to let the model find them for you, however, and a number of options are provided to help you fine-tune the result.
- Rules are applied in the order listed, with the first matching rule determining the outcome for a given record. Taken independently, rules or conditions may overlap, but the order of rules resolves ambiguity. If no rule matches, the record is assigned to the remainder rule.

To obtain a predictive rule model

1. Specify a data source. This can be any analytical source that records the outcome you want to predict.

2. Specify a target field that records the result you want to predict.

3. Specify the specific value or response you want to find. For example, if the target field is named *response*, you could choose *True* or *False* to find customers who responded--or who didn't--accordingly.

   For example, the target might indicate which customers have churned, responded to past offers, raised fraudulent claims, etc. All other fields will automatically be included as inputs that may be useful in predicting this value.

4. Specify optional settings as desired. See the topic "Optional model settings" on page 44 for more information.

5. If desired, click the Data Overview icon to see an overview of the data that will be used to build the current model. See the topic "Data overview" on page 24 for more information.

6. Click **Start Build**.

7. Specify whether you want to find segments with high or low probability relative to the specified target. For example, you could look for groups or customers who are mostly likely to churn, or least likely, and include them accordingly.

8. Specify the maximum number of segments to be found. Typically this number is kept small in the interest of simplicity and performance.

9. Click **Grow Model** to obtain a list of one or more segment rules that identify groups with higher or lower probability.

10. Click the **Include**/**Exclude** icons to specify how the rules are to be used and how the relevant segments are scored.

    Included segments are given a score that matches the target value (for example *response=true*). Excluded segments are scored as *NULL,* but are not removed from the file.

11. Optionally, you can manually add or edit segments as desired. See the topic "Defining segment rules" on page 31 for more information.

12. Optionally, use the Evaluate and Test features to see how the model performs on your sample data.

13. Save the model before closing the model builder or returning to the application.

14. Click **Use Model**, and select the model field you want to use. For example, if you want to use the value predicted by the model as input to a rule, select the field that contains the predictions.

Tips:

- Use the up/down arrows to move rules up or down. Each record is scored based on the first hit, so the order of rules will change the result. For example, a record that matches multiple rules will be scored based on the first that applies.

- Change one or more options and click **Grow Model** again to find additional rules. For example, you might lower the threshold for minimum segment or confidence interval in order to identify additional segments.

- Clicking **Grow Model** does not clear the table or rebuild the model from scratch, it only finds more segment rules to add to the table.

Building a clustering model

Clustering models allow you to categorize records into a certain number of clusters. This can help you identify natural groups in your data.

Clustering models focus on identifying groups of similar records and labeling the records according to the group to which they belong. This is done without the benefit of prior knowledge about the groups and their characteristics. In fact, you may not even know exactly how many groups to look for. This is what distinguishes clustering models from the other machine-learning techniques—there is no predefined

output or target field for the model to predict. These models are often referred to as **unsupervised learning** models, since there is no external standard by which to judge the model's classification performance. There are no *right* or *wrong* answers for these models. Their value is determined by their ability to capture interesting groupings in the data and provide useful descriptions of those groupings.

Clustering methods are based on measuring distances between records and between clusters. Records are assigned to clusters in a way that tends to minimize the distance between records belonging to the same cluster.

Clustering models are often used to create clusters or segments that are then used as inputs in subsequent analyses. A common example of this is the market segments used by marketers to partition their overall market into homogeneous subgroups. Each segment has special characteristics that affect the success of marketing efforts targeted toward it. If you are using data mining to optimize your marketing strategy, you can usually improve your model significantly by identifying the appropriate segments and using that segment information in your predictive models.

To obtain a clustering model

1. Specify a data source. This can be any analytical source that records the outcome you want to predict.
2. Specify optional settings as desired. See the topic "Optional model settings" on page 44 for more information.
3. If desired, click the Data Overview icon to see an overview of the data that will be used to build the current model. See the topic "Data overview" on page 24 for more information.
4. Click **Find Clusters**.
5. Optionally, you can add manual clusters. See the topic "Using manual clusters" for more information.
6. Optionally, use the Evaluate and Test features to see how the model performs on your sample data.
7. Save the model before closing the model builder or returning to the application.
8. Click **Use Model**, and select the model field you want to use. For example, if you want to use the value predicted by the model as input to a rule, select the field that contains the predictions.

Using manual clusters

If desired, you can manually define clusters, then use automated clustering to find clusters of records not identified by any of the manually-defined clusters. Create manual clusters via a "first hit" segment-based allocation rule set. The rules you create will define clusters and the remainder will be used for auto cluster modeling.

Note that manual clusters are not included in results summary charts.

1. In the **Manual Cluster** section, click the **Create a new rule icon** and add one or more rules to define manual clusters.
2. If desired, change cluster label names.
3. Click the **Records in manual clusters** icon to perform a record count of the total number of records included in the manual clusters. The **Count** column will be updated for each manual cluster defined.

Building an association model

Association models allow you to predict which items are most likely to appear together, and predict the strength of the relationship between them.

Association rules associate a particular conclusion (the purchase of a particular product, for example) with a set of conditions (the purchase of several other products, for example). For example, the rule

```
beer <= cannedveg & frozenmeal (173, 17.0%, 0.84)
```

states that *beer* often occurs when *cannedveg* and *frozenmeal* occur together. The rule is 84% reliable and applies to 17% of the data, or 173 records. Association rule algorithms automatically find the associations that you could find manually using visualization techniques, such as the Web node.

The advantage of association rule algorithms over the more standard decision tree algorithms (C5.0 and C&R Trees) is that associations can exist between *any* of the attributes. A decision tree algorithm will build rules with only a single conclusion, whereas association algorithms attempt to find many rules, each of which may have a different conclusion.

The disadvantage of association algorithms is that they are trying to find patterns within a potentially very large search space and, hence, can require much more time to run than a decision tree algorithm. The algorithms use a **generate and test** method for finding rules--simple rules are generated initially, and these are validated against the dataset. The good rules are stored and all rules, subject to various constraints, are then specialized. **Specialization** is the process of adding conditions to a rule. These new rules are then validated against the data, and the process iteratively stores the best or most interesting rules found. The user usually supplies some limit to the possible number of antecedents to allow in a rule, and various techniques based on information theory or efficient indexing schemes are used to reduce the potentially large search space.

To obtain an association model

1. Select a data source. This can be any analytical source that records the outcome you want to predict.
2. Specify optional settings as desired. See the topic "Optional model settings" for more information.
3. If desired, click the Data Overview icon to see an overview of the data that will be used to build the current model. See the topic "Data overview" on page 24 for more information.
4. Click **Build Model**.
5. Optionally, use the Test features to see how the model performs on your sample data. Note that Evaluation is not available for the Association model type.
6. Save the model before closing the model builder or returning to the application.
7. Click **Use Model**, and select the model field you want to use. For example, if you want to use the value predicted by the model as input to a rule, select the field that contains the predictions.

With other model types, you can choose between different modeling techniques. For association models, the *Apriori* model type is used.

Optional model settings

When building models, you can choose from the following optional settings. Note that depending on the type of model and the target selected, you may not see all of these settings. If changing any of these settings for an existing model, the model must be rebuilt for changes to apply.

Automatically clean up and prepare data for reliable model building. Identifies and repairs data issues to make modeling faster, more predictable, and more reliable. Screens fields that are problematic or not likely to be useful, for example by handling missing and extreme values, deriving new attributes when appropriate, and improving performance through intelligent screening and sampling techniques. The first time a model is built with a new data source, the analysis is done to identify issues and fixes, which may slow performance on the first pass only. On subsequent runs, the fixes are applied, but the analysis is not repeated unless the data source changes. This setting may be disabled for some models, including those with custom data preparation settings specified by an expert user.

Note: Automatic data preparation settings only apply when creating a new model, and binning is not performed.

Automatically partition data to enable model evaluation on build data source for evaluation and testing. Selecting this option splits the data into separate subsets or samples for training and testing the model. By building the model on one subset and testing it on another, you can get an idea how it will generalize to other data sets. You can also specify a percentage value for the randomly generated *training partition size* and *testing partition size*.

The minimum training partition size is 1 and the maximum 100. The minimum testing partition size is 0 and the maximum is 100. The minimum validation partition size is 0 and the maximum is 100. The total percentage must be 100 or less for the model to build successfully.

These partitioning options are available for all model types except association modeling.

Choose model techniques for model building. If desired, click **Select** to choose which modeling techniques to include when building the model. Based on the target you selected, all available modeling techniques will be listed. The modeling techniques available (and those selected by default) will vary depending on the target you selected. For example, the *Decision List* algorithm will only appear for *flag* targets. This feature is only available for predictive models.

For complete details about the modeling techniques, see the *IBM SPSS Modeler Algorithms Guide* and other documentation shipped with the IBM SPSS Modeler product and available on the Web.

Maximum # of models to be combined. Allows you to set the maximum number of models to be retained and combined. If you set this option to 1, a single model will be built and retained. If there are less models built than the value entered here, then all models built will be retained and combined. This option is only available for predictive models.

Profit Criteria (used to rank models with binary targets). Allows you to set profit criteria prior to building the model. The values set here will be set as the default for evaluating the model. This option is only available for predictive models with a flag target selected.

Specify inputs to use. Allows you to select the fields you want to use. Typically these would be those that have some practical relationship to the thing you are trying to predict, such as age or income. If you have a large data set, limiting the number of fields is one way of simplifying the model. If the data contains fields such as customer id or contact number, these would not typically be useful in modeling and should not be selected. Fields that duplicate other data may also be excluded.

Clicking on a linked input field (an expression) opens the expression viewer for that expression. To edit an expression, see the Data tab. See the topic "Expression Builder" on page 26 for more information.

Specify selections to use. Specifies which records to include or exclude when modeling. You can search for existing rules, or create new ones as appropriate. See the topic "Defining selection rules" on page 33 for more information.

In addition, if global selections have been defined, they will be displayed here, and you can specify whether they should also apply during modeling. You can either choose to use all or none of the global selection rules; you cannot choose a subset.

Segmentation options

When building predictive rule models, a number of segmentation options are available. These are in addition to the options available with other types of models.

Minimum segment size. These settings dictate the minimum segment size, with the larger of the two values taking precedence. For example, if the percentage value equates to a number higher than the absolute value, the percentage setting takes precedence.

- **Percentage of previous segment.** Specifies the minimum group size as a percentage of records. The minimum allowed setting is 0; the maximum allowed setting is 99.9.
- **Absolute value.** Specifies the minimum group size as an absolute number of records. The minimum allowed setting is 1; there is no maximum setting.

Maximum number of attributes. Specifies the maximum number of conditions per segment rule. The minimum allowed setting is 1; there is no maximum setting.

Allow attribute re-use. When enabled, each cycle can consider all attributes, even those that have been used in previous cycles. The conditions for a segment are built up in cycles, where each cycle adds a new condition. The number of cycles is defined using the **Maximum number of attributes** setting.

Confidence interval for new conditions. Specifies the confidence level for testing segment significance. This setting plays a significant role in the number of segments (if any) that are returned as well as the number-of-conditions-per-segment rule. The higher the value, the smaller the returned result set. The minimum allowed setting is 50; the maximum allowed setting is 99.9.

Auto cluster options

When building clustering models, a number of auto cluster options are available. These are in addition to the options available with other types of models.

Evaluation field. Specifies the field to use for evaluation. If an evaluation field is selected, the ranking method in the IBM SPSS Modeler auto cluster build node will be set to *Importance of evaluation*. If no evaluation field is selected, the ranking method will be left as the default. Note that an analytical field can be selected as the evaluation field here.

Set a desired range for the number of clusters found. Specifies the desired minimum and maximum range for the number of clusters found by the auto cluster model.

Association rule and scoring options

When building association models, a number of association rule and scoring options are available. These are in addition to some of the options available with other types of models.

Association Rule Options.

- **Coverage (%).** Specify the minimum coverage percentage to control how generally applicable the rules will be. For example, setting a value of 80% means that all rules will apply to at least 80% of cases. Note that higher values will yield fewer rules, or potentially none at all.
- **Confidence (%).** Specify the minimum confidence percentage.
- **Maximum number of preconditions.** Specify the maximum number of preconditions.
- **Data are already sorted by ID (transactional format only).** Select this option to disable the sort that takes place by default at the start of model building to ensure IDs are contiguous. Selecting this option will reduce processing time. This option has no impact if the data format is tabular.

Scoring Options.

- **Maximum number of predictions.** Specify the maximum number of predictions included. This setting is used in conjunction with the rule criterion drop-down to produce the "top" predictions, where top indicates the highest level of confidence, support, lift, and so on.
- **Rule criterion.** Select the measure used to determine the strength of rules. Rules are sorted by the strength of criteria selected here in order to return the top predictions for an item set. Available criteria are:
  - **Confidence** displays the ratio of rule support to rule condition support. This indicates the proportion of records with the specified rule condistion(s) for which the rule prediction(s) is/are also true. For example, if 50% of the training data contains bread (indicating rule condition support) but only 20% contains both bread and cheese (indicating rule support), then the prediction for the rule `bread -> cheese` would be `Rule Support / Rule Condition Support` or, in this case, 40%.
  - **Coverage (%)** displays rule condition support--that is, the proportion of IDs for which the rule conditions are true, based on the training data. For example, if 50% of the training data includes the purchase of bread, then the `rule bread -> cheese` will have a rule condition support of 50%.
  - **Rule Support (%)** displays the proportion of records for which the entire rule, rule condition(s), and rule prediction(s), are true. For example, if 20% of the training data contains both bread and cheese, then rule support for the rule `bread -> cheese` is 20%.

- **Lift** displays the ratio of confidence for the rule to the prior probability of having the rule prediction. For example, if 10% of the entire population buys bread, then a rule that predicts whether people will buy bread with 20% confidence will have a lift of 20/10 = 2. If another rule tells you that people will buy bread with 11% confidence, then the rule has a lift of close to 1, meaning that having the rule condition(s) does not make a lot of difference in the probability of having the rule prediction. In general, rules with lift different from 1 will be more interesting than rules with lift close to 1.

- **Deployability (%)** is a measure of what percentage of the training data satisfies the rule condition(s) but does not satisfy the rule prediction. In product purchase terms, it basically means what percentage of the total customer base owns (or has purchased) the condition(s) but has not yet purchased the rule prediction(s). The deployability statistic is defined as `((Rule Condition Support in # of Records - Rule Support in # of Records) / Number of Records) * 100`, where `Rule Condition Support` means the number of records for which the rule conditions are true and `Rule Support` means the number of records for which both rule conditions and the prediction are true.

- **Check entire transaction.** Select from the following options:
  - **Ensure predictions not present** ensures that rule predictions are not also present. For example, if the purpose of scoring is to make a home furniture product recommendation, then it is unlikely that a basket that already contains a dining room table will be likely to purchase another one. In such a case, you should select this option. On the other hand, if products are perishable or disposable (such as cheese, baby formula, or tissue), then rules where the consequent is already present in the basket may be of value. In the latter case, the most useful option might be **Do not check for predictions**.
  - **Ensure predictions are present** ensures that rule predictions are also present in the basket. This approach is useful when you are attempting to gain insight into existing customers or transactions. For example, you may want to identify rules with the highest lift and then explore which customers fit these rules.
  - **Do not check for predictions** includes all rules when scoring, regardless of the presence or absence of rule predictions in the basket.

- **Allow repeat predictions.** Select this option to include multiple rules with the same rule prediction when scoring. For example, selecting this option would allow the following rules to be scored:

  ```
  bread & cheese -> wine
  cheese & fruit -> wine
  ```

  Turn off this option to exclude repeat predictions when scoring.

- **Ignore unmatched items.** Select this option to ignore the presence of additional items in the item set. For example, when this option is selected for a basket that contains [`tent & sleeping bag & kettle`], the rule `tent & sleeping bag -> gas stove` will apply despite the extra item (`kettle`) present in the basket.

Evaluating models

After building or opening a completed model, click **Evaluate** to obtain charts such as distribution, error distribution, coincidence matrix, gains, lift, response, profit, and ROI, and to simulate the maximum profit possible. The chart types and other options available will vary depending which model type was selected. The Evaluate dialog is only available if the selected model is complete. Note that evaluation is not available for the Association model type.

1. In the upper panel, select an evaluation data set. As with model building, this must be an analytical data set for which actual responses are known. (Otherwise you have no way to compare actual responses to those predicted by the model.)

2. If you want to modify the models to include in the evaluation, click **Modify Model Selection**. After making selections, you can also click the following **Copy current model selections back to individual model results table** button, if desired. These options are only available for the Predictive model type.

Figure 6. Copy current model selections back to individual model results table

3. Select the graphs you want to build.

4. If simulating profit, select the **Simulate maximum profit** box and enter the population on which to base the profit calculation, the budget, and the minimum value. See the topic "Simulating maximum profit" on page 49 for more information.

5. If requesting a profit or ROI chart, or if simulating profit, specify cost and revenue values. These values indicate the cost for each offer, and the amount of revenue you expect to make from each favorable response, respectively.

6. Select the response value that is considered as a "hit." This must be one of the values for the target field used to build the model, such as *response = true* to indicate a positive or favorable response.

7. If requesting a performance chart, you can also select the **Include individual model evaluations for performance charts** option.

8. If desired, select **Include tabular results**. This requires more time for the evaluation results to be returned.

9. Click **Run** to view the results.

Distribution charts (all targets)

Distribution charts plot observed versus expected values, and are displayed by default when the model completes. The Evaluate dialog box allows you to obtain distribution charts for a data set other than the testing data set. For range targets, the distribution chart is the only chart type available.

You can choose from the following additional charts:

Error Distribution charts

The error distribution charts displays a histogram of the errors in the model outcomes, where the error is the difference between the predicted value and the actual value. A positive error arises when the predicted value is greater than the actual value, and a negative error arises when it is less. This chart is only available for continuous targets with the Predictive model type.

Coincidence matrix charts

A coincidence matrix plots the counts of *predicted* versus *actual* outcomes. For a numeric target, the range is divided into the number of bins you specified.

Gains charts

Gains are used to show the gain or "lift" that can be obtained by using the model, and are defined as the proportion of hits in each increment relative to the total number of available hits. The flat diagonal line plots the expected response for the entire sample, if no model were used. In this case, the response rate would be constant, since one person is just as likely to respond as another. To double your yield, you would need to ask twice as many people. The curved line indicates how much you can improve this result by targeting those who are most likely to respond. The steeper the curve, the higher the gain.

Lift charts

The lift chart plots the percentage of records in each increment that are hits with the overall percentage of hits in the training data set.

Response charts

The response chart plots the percentage of records in the increment that are hits.

Profit charts

Profit equals the revenue for each record minus the cost for the record. Profits for a quantile are simply the sum of profits for all records in the quantile. Revenues are assumed to apply only to hits, but costs apply to all records. See the topic "Simulating maximum profit" for more information.

ROI charts

ROI (return on investment) is similar to profit in that it involves defining revenues and costs. ROI compares profits to costs for each interval.

Simulating maximum profit

Based on the response rates predicted by the model, you can simulate the maximum profit to be made for a given population. To do this you need to specify cost and revenue values, as well as the size of the population.

To predict the maximum profit:

1. Specify the population, or total number of records, on which the prediction should be based. *Note*: When scoring a model, this is automatically set to the total number of records in the data source or subset.
2. Specify a value for budget if you want the budget parameter to be included in the profit simulation, and a minimum value.
3. Specify the cost associated with each record. For example, this could be the print and mailing cost, per person, of promotional literature for a campaign.
4. Specify the expected revenue that a positive response would generate. For example, the retail cost of an item in a campaign.
5. Select the response you are interested in.
6. Click **Run**.

The predicted results graph shows the best result for maximum profit, and identifies the percentage of records you would need to target to achieve this, based on propensity to respond as identified by the model.

Beneath the graph you can move a slider bar from zero to 100% of the records to target to see how this would affect your profit.

Profit comparison

To verify that the chosen records are a good selection, you can compare the likely profit to be made from those records, a random group of the same percentage of records, and the whole set of records.

Details shown include the number of records targeted, the likely response rate, costs, income, and overall profit. You can export this information in HTML format for use in other applications.

Profit chart

You can display the predicted cumulative or non-cumulative profit from the selected records as a graph.

Note: This graph is generated when you first run the evaluation and is not automatically updated with any changes you make via the slider in the profit simulation section.

Using model scores in applications

To score a model means to apply it to some data or population in order to obtain a result or prediction that can be used as input to decisions. For example, you might score all customers in your database to identify prospects for a direct mail campaign, or score customers one at time as they contact your call center in order to decide which products or services to offer them. Depending on the application, scoring results can be written to a database table or flat file, or used as inputs to the segment, selection, and allocation rules that drive decisions in an application.

Model scores can be used in the following ways:
- As input to rules. See the topic "Creating rules" on page 31 for more information.

Model output fields

By convention, the names of the scoring output fields are based on the name of the target field, but with a custom prefix added. For example if the target field is named *Response*, the output fields will have names like *$XF-Response* and *$XFRP-Response*. The specific fields depend on the type of model, and the measurement level for the selected target field, as detailed below. See the topic "Measurement levels" on page 24 for more information.

Flag targets. Automated models with flag targets output the following fields:
- **$XF-<target>.** Predicted value for each record, reported in terms of the "true" and "false" values defined for the field.
- **$XFRP-<target>.** Propensity score; indicates the likelihood of the "true" value for each record. Propensity scores differ from confidence scores, which are typically reported in terms of the prediction for that record. For example a "false" prediction with a high confidence will translate to a low propensity as this would indicate a high likelihood *not* to respond.

Categorical targets. Automated models with categorical targets (nominal or ordinal) output the following fields:
- **$XS-<target>.** Predicted value for each record.
- **$XSC-<target>.** Confidence value associated with the prediction

Continuous targets. Automated models with continuous numeric targets output the following fields:
- **$XR-<target>.** Predicted value for each record.
- **$XRE-<target>.** Standard error for the prediction.

Predictive rule (interactive) models. Predictive rule models (flag or categorical targets only) output the following fields:
- **$D-<target>.** Predicted outcome for each record, with the "true" value defined for the field used to indicate the desired response, and null indicating any other value.
- **$DP-<target>.** Probability of desired outcome for each record; matches the probability displayed for the segment in the results table when building the model. See the topic "Building a predictive rule model" on page 41 for more information.
- **$DI-<target>.** Index number used to identify the segment into which the record falls (displayed in the left-most column in the modeling results).
- **$DRP-<target>.** Propensity score indicating the likelihood of the desired response. Available for flag targets only.

Chapter 6. Administering applications

Administrators can control the way business users interact with IBM Analytical Decision Management applications. This level of control makes business users' jobs easier so they can more quickly and efficiently use the application to help solve business problems. It also enables administrators to restrict access to certain functionality.

Use IBM SPSS Collaboration and Deployment Services Deployment Manager to grant administrative authority to specific users or groups as needed. Only users who are granted the *Decision Management Administration* action have access to the features described in this section.

1. Open IBM SPSS Collaboration and Deployment Services Deployment Manager.
2. From the Tools menu, select **Server Administration**.
3. On the Server Administration tab, connect to your server.
4. Under **Users and Groups**, select **Local User Repository**.
5. Click **New Group** to create a new group for administrators and a new group for business users. For example:
   - Decision Management Administrators
   - Decision Management Users

   You can also create new users at this time, if desired. While creating the groups, add the appropriate users to each group. For example, members of the administrators group will be granted authority to administer IBM Analytical Decision Management applications at your organization.
6. Go to Roles and click **New Role** to create a new administrative role and a new business user role. For example:
   - IBM Analytical Decision Management administrator
   - IBM Analytical Decision Management user

   While creating the new roles, assign them the following actions, at a minimum.

Table 6. Role actions.

Role	Actions Required
Decision Management administrator	Access Contents and Folders
	Decision Management Administration
	Job Edit
	Run Report Dynamically
	Score Model
	Submit Work
	Show All Versions
Decision Management user	Access Contents and Folders
	Job Edit
	Run Report Dynamically
	Score Model
	Submit Work
	Show latest *OR* Show All Versions*

Note: Either the *Show latest* or *Show All Versions* action is required if users need access to project versions they did not create. Otherwise projects will open in read-only mode if the current user does not have access to the latest version.

7. Select the Decision Management administrator role and click **Edit Users and Groups**. Add the Decision Management Administrators group you just created.
8. Select the Decision Management user role and click **Edit Users and Groups**. Add the Decision Management Users group you just created.

General options

Certain general options are available to administrators on all tabs of the IBM Analytical Decision Management user interface. These options allow administrators to hide or lock tabs from business users.

- The top of each tab has the option **Hide <tab name> tab from non-administrators**, where **<tab name>** is the name of a tab. This hides the entire tab from business users. For example, if you don't want users to see the Data tab for an application, log on to the application, go to the Data tab, and select **Hide Data tab from non-administrators**.
- The top of each tab has the option **Lock all <tab name> options**, where **<tab name>** is the name of the tab. This locks all controls on the tab so they're read-only for the business user. For example, if you want business users to be able to view current data settings but not change them, you can set this option on the Data tab of an application. This setting is also very useful for situations when you want to set certain things up in advance for the business user to simplify their job.
- Administrators can also lock and unlock individual settings on each tab by clicking the padlock icon when available.

For example, in applications with a Define tab, administrators can lock the dimension hierarchy tree so business users cannot change it. Clicking the padlock icon locks or unlocks its associated control. Business users will see the locked padlock icon for controls the administrator locks, but will never see an unlocked padlock icon.

Individual items that are locked may also be displayed with a grayed-out background or muted colors to indicate that they are unavailable for selection; an error message is displayed if a user attempts to select them. When an entire tab is locked a message to that effect is displayed at the top of the tab.

Controlling application shortcuts

The Applications Launch Page provides an easy way for business users to open applications and add custom shortcuts to specific, labeled versions of applications. Administrators can specify which application shortcuts will show up on the launch page for all users. After logging on with administrator authority:

- If desired, click the padlock icon on an application shortcut so business users cannot remove it from their launch page. Note that business users can still add and remove other applications shortcuts. Users will see the locked padlock icon instead of the close icon for shortcuts that are locked. See the topic "Launching applications" on page 9 for more information.

Hiding the Home page

Each application provides a step-by-step workflow, as represented by the graphics on the Home page. Clicking a graphic jumps to that section of the user interface. However, business users may not always need access to every step in an application's workflow.

Administrators can hide the Home page from business users. For example, in a certain application business users might only use the Reports tab. In this case an administrator could hide the Home page and all other tabs from business users. So when the business user logs on he is taken directly to the Reports tab and never sees any other tabs.

1. On the Home page, select **Hide Application home page**.
2. In the drop-down, specify which tab to display when users first open the application.

   Note that if you subsequently hide this selected tab, the next unhidden tab in the application is displayed.

Locking data options

Data

The Data tab defines the data sets the application uses for analysis, simulation and testing, scoring, or other operational use. These data sources are defined relative to your server installation; hence all paths to data files, ODBC sources, and so on are all relative to the server.

As an administrator, you may want to set up some or all data before business users start using their applications. Administrators can:

- Select **Lock Datasource** when creating a new data source to lock it after it's created.
- Lock and unlock individual data sources by clicking the padlock icon next to a data source in the table, or in the Data Source Editor dialog.
- Set and lock the data model used in the application by selecting a data model in the **Project Data Model** drop-down and clicking the padlock icon.

  Note that business users can copy a locked data source from Project Data Sets to their My Data Sets section, but it remains locked in both sets. See the topic Chapter 3, "Managing data sources," on page 15 for more information.
- If the project data model is a *Data View*, use the **Allow use of ODM rules based on hierarchical data** drop-down to control the usage of IBM Operational Decision Management (ODM) rules that have a hierarchical data model. Use of such rules can result in poor performance. The following options are available in the drop-down:
  - **Don't allow** is recommended for best performance. Disallows ODM rules with a complex data structure from being used anywhere in the project. An error is displayed if a user selects a complex rule at any point in the application.

Downloading project metadata for use in creating external rules

Rules created in a Business Rules Management System such as IBM Operational Decision Management can be referenced and used in IBM Analytical Decision Management applications, provided they have been developed to support the same data model used in the current IBM Analytical Decision Management project, and can be deployed as a Web service for use by the application. To accomplish this, an IBM Analytical Decision Management administrator must download the relevant metadata. There are two ways to download the metadata, and the complexity of the data model and the potential resultant rule complexity differ accordingly:

Downloading project metadata from IBM Analytical Decision Management

An administrator can click the **Download metadata** icon in IBM Analytical Decision Management to download a *.ZIP* file containing metadata for the current project. The downloaded file is saved locally on the system used by the administrator to trigger the download.

Figure 7. Download metadata icon

The metadata downloaded this way supports the creation of rules based on the project data model. The project data model will be flat (not hierarchical), so any rules authored based on this metadata can only

utilize a flat data model. A rule authored based on this flat data model may be referenced in an IBM Analytical Decision Management project no matter the type of the project data model data source.

The following files are included in the *.ZIP* file:

- **XML schema definition (*.XSD).** Contains definitions of the field types in the data model for the current IBM Analytical Decision Management project. This file can be imported into an external development tool such as IBM Operational Decision Management Rule Designer in order to develop rules for use with the current project.
- **Web Service Description Language template (*.WSDL).** Included to assist in development of Web services that use the data model. The *.WSDL* file is provided as a template that can be used to build Web services responsible for mediating messages between the IBM Analytical Decision Management application and an external rule service. Details on the mediating Web service can be manually supplied by the integrator.

Downloading data model metadata for an IBM SPSS Collaboration and Deployment Services analytic data view

In IBM SPSS Collaboration and Deployment Services Deployment Manager, you can export metadata for a particular analytic data view.

Metadata downloaded in this way supports the creation of rules based on the logical data model of an analytic data view. An analytic data view's logical data model may be hierarchical. An IBM Analytical Decision Management Data View source supports generation of an XML representation of the hierarchical data for each record, which can then be passed to external rules during execution. Rules authored based on this metadata may be complex and utilize the hierarchical data model.

In IBM Analytical Decision Management, an external rule based on a hierarchical data model may only be referenced in a project where the project data model uses a Data View source. To ensure compatibility, the Data View source and the external rule must be based on the same analytic data view's logical data model. Only the Data View source can generate the hierarchical data based on the analytic data view instance it references. In other words, to use external rules, the project data model in the IBM Analytical Decision Management project needs to match that used to create the rule, and the analytic data view makes that possible.

To download a *.ZIP* file with metadata for the analytic data view's logical data model:

1. In IBM SPSS Collaboration and Deployment Services Deployment Manager, open an analytic data view from the repository and go to the **Logical** tab.
2. Right-click a table and export the XOM metadata.
3. Select a location to save the file to and click **Save**.

The following files are included in the *.ZIP* file:

XML schema definitions (*.XSD). Two or more XML schema definitions are included. One contains a definition which may be useful for a rule response type. The other schema(s) contain the definitions of the tables, including hierarchical relationship detail and field types in the logical data model of the analytic data view.

You can import these schema files into an external development tool such as IBM Operational Decision Management Rule Designer in order to develop rules for use with a project whose project data model uses a Data View source based on the same analytic data view instance.

For information on creating external rules for use in IBM Analytical Decision Management, see the *Application Designers Guide*.

Note that administrators may restrict use of these complex external rules because they can degrade performance. For details, see "Locking data options" on page 53.

Editing the XML schema definition

The .*XSD* file downloaded from IBM Analytical Decision Management contains two complex data types: one representing the IBM Analytical Decision Management project data model, which should be used for the input parameter of IBM Operational Decision Management rules projects; and one representing a rules return type, which should be used for the output parameter of IBM Operational Decision Management rules projects.

The IBM SPSS Collaboration and Deployment Services Deployment Manager analytic data view metadata contains two or more.*XSD* files: one relating to the logical data model of an analytic data view, which should be used for the input parameter of IBM Operational Decision Management rules projects; and one or more containing a complex type representing a rules return type, which should be used for the output parameter of IBM Operational Decision Management rules projects.

Expert users may choose to edit the generated .*XSD* schema definition to tailor them for the specific rule or rules they wish to create. There are circumstances when the user wants to edit the input and/or the response related types defined in the related .*XSD* schema definition prior to creating their rules. For instance, the expert user may need to edit the input parameter related schema to permit their authored rules to handle the potential of input fields with null values. The expert user can do this by indicating via the .*XSD* schema definition which input parameter fields/elements are required and which are optional.

Editing the input parameter type

The number of possible occurrences of each field-related element is not explicitly set in the .*XSD* schema (i.e., the attributes `minOccurs` and `maxOccurs` are not being set, and the default value for both is 1). As such, if the schema is not edited before using it to generate IBM Operational Decision Management rules, all fields in the input parameter type will be required by the generated rules.

Editing the project data model related input parameter type

If the schema defining the input parameter type was generated by downloading an IBM Analytical Decision Management project's metadata, the input parameter's data model will be flat. In this case, IBM Analytical Decision Management will directly create the payload required by the rule service on execution. This gives the expert user more freedom when editing the input parameter type and still maintaining compatibility (i.e., the expert user can remove fields/elements they will not require for rule authoring, or indicate fields/elements are optional).

Editing the analytic data view logical data model related input parameter type

If the schema defining the input parameter type was generated by exporting an analytic data view's logical data model from IBM SPSS Collaboration and Deployment Services Deployment Manager client, the input parameter's data model may be hierarchical. If it is hierarchical, then IBM Analytical Decision Management will *not* directly create the payload required by the rule service on execution. The main part of the payload will be generated by the Analytic Data View source. This limits how the expert user may edit the input parameter type and maintain compatibility (for example, the expert user can only indicate fields/elements are optional).

General information

Prior to invoking a rule created in a Business Rules Management System such as IBM Operational Decision Management, IBM Analytical Decision Management asks the rule's service whether its input parameter data model is hierarchical or flat. If the input parameter is flat, it also checks which inputs are required by the rule's service.

If the input parameter's data model is flat, and the record to score contains a null value for any of the rule service's required fields, IBM Analytical Decision Management will not invoke the rule, and the output/response fields of the rule are populated with default values. The value depends on the type of the output/response field as follows:

Table 7. Values for response field types.

Response field type	Value
xsd:string	""
xsd:boolean	""
xsd:int	0
xsd:long	0
xsd:byte	0
xsd:double	0.0
xsd:float	0.0
xsd:date	0.0
xsd:time	0.0
xsd:datetime	0.0

This can impact the results of scoring records via a rule service, and care should be taken in editing the input parameter in the *.XSD* schema.

If the input parameter's data model is hierarchical, and the record to be scored contains a null value for any of the rule service's required fields, IBM Analytical Decision Management will invoke the rule, but the field or fields with a null value will be omitted from the payload created by the Data View source node. If the omitted field was a required field, this will cause the rule invocation to fail, so again care should be taken when editing the input parameter in the *.XSD* schema.

Editing the schema requires knowledge of the rule services that will be created. If a particular input field is not required by the rule, and the rule's input parameter data model is **flat**, then the field-related element may either be **removed** from the input parameter in the schema or **marked as optional**. If a particular field is not required by the rule, and the rule's input parameter is **hierarchical**, the field-related element **cannot be removed**; but it may be **marked as optional**. This is because IBM Analytical Decision Management does not directly generate the payload for rules based on a hierarchical data model.

When you remove an input parameter field, the field value is ignored when the rule is called and the rule cannot refer to the field. When you make an input parameter field optional, a null value is allowed and the rule must be written defensively to check for that. For example, if the rule's input parameter data model is flat and you know the field Age will not be used by the rule, you can remove the related element to indicate it is not a required field. In the following *.XSD* schema example, if the field Age is not required, the entire section <xsd:element name="Age"> can be removed from the schema:

```
<xsd:complexType name="drugs">
<xsd:annotation>
  <xsd:appinfo>
   <dmname>drugs</dmname>
  </xsd:appinfo>
 </xsd:annotation>
<xsd:sequence>
  <xsd:element name="Age">
   <xsd:annotation>
    <xsd:appinfo>
     <dmname>Age</dmname>
    </xsd:appinfo>
   </xsd:annotation>
   <xsd:simpleType>
```

```
    <xsd:restriction base="xsd:long">
     <xsd:minInclusive value="15"/>
     <xsd:maxInclusive value="74"/>
    </xsd:restriction>
   </xsd:simpleType>
  </xsd:element>
  <xsd:element name="Sex">
   <xsd:annotation>
    <xsd:appinfo>
     <dmname>Sex</dmname>
    </xsd:appinfo>
   </xsd:annotation>
   <xsd:simpleType>
    <xsd:restriction base="xsd:string">
     <xsd:enumeration value="F"/>
     <xsd:enumeration value="M"/>
    </xsd:restriction>
   </xsd:simpleType>
  </xsd:element>
  ...
 </xsd:sequence>
</xsd:complexType>
```

Or if the rule's input parameter data model is flat or hierarchical, you might want to make an input field element optional if it's not required by the rule(s)--by setting the element's minOccurs attribute value to 0:

```
<xsd:complexType name="drugs">
 <xsd:annotation>
  <xsd:appinfo>
   <dmname>drugs</dmname>
  </xsd:appinfo>
 </xsd:annotation>
 <xsd:sequence>
  <xsd:element name="Age" minOccurs="0">
   <xsd:annotation>
    <xsd:appinfo>
     <dmname>Age</dmname>
    </xsd:appinfo>
   </xsd:annotation>
   <xsd:simpleType>
    <xsd:restriction base="xsd:long">
     <xsd:minInclusive value="15"/>
     <xsd:maxInclusive value="74"/>
    </xsd:restriction>
   </xsd:simpleType>
  </xsd:element>
  <xsd:element name="Sex">
   <xsd:annotation>
    <xsd:appinfo>
     <dmname>Sex</dmname>
    </xsd:appinfo>
   </xsd:annotation>
   <xsd:simpleType>
    <xsd:restriction base="xsd:string">
     <xsd:enumeration value="F"/>
     <xsd:enumeration value="M"/>
    </xsd:restriction>
   </xsd:simpleType>
  </xsd:element>
  ...
 </xsd:sequence>
</xsd:complexType>
```

Hierarchical data model example:

```
<xsd:complexType name="Customer">
 <xsd:sequence>
  <xsd:element name="FirstName" type="xsd:string" minOccurs="0"/>
  <xsd:element name="Sirname" type="xsd:string" minOccurs="0"/>
  <xsd:element name="Age" type="xsd:long"/>
  <xsd:element name="Sex" type="xsd:string"/>
```

```
  <xsd:element name="CustID" type="xsd:string"/>
  <xsd:element maxOccurs="unbounded" minOccurs="0" name="Orders" type="xom:Order"/>
 </xsd:sequence>
</xsd:complexType>
<xsd:complexType name="Order">
 <xsd:sequence>
  <xsd:element name="CustID" type="xsd:string"/>
  <xsd:element name="NoOfItems" type="xsd:long"/>
  <xsd:element name="TotalCost" type="xsd:long" minOccurs="0"/>
  <xsd:element name="Date" type="xsd:date"/>
 </xsd:sequence>
</xsd:complexType>
```

The rules can also be written defensively in order to handle null values, in which case the field can be marked as `optional` in the schema. Following is an example of an IBM Operational Decision Management rule called *currentOfferIsReplacePhone* that has been written to deal with null values:

```
if
    the CURRENT OFFER of 'the customer' is not null
    and the CURRENT OFFER of 'the customer' is "Replace Phone"
then
    make it true that 'the response' is value ;
else
    make it false that 'the response' is value ;
```

Editing the response parameter type

Note that it is possible to create and utilize IBM Operational Decision Management rules that return a variety of types to IBM Analytical Decision Management (*boolean* for use in Selections, *numeric* for use in aggregate rules for example, and *string*). However, to create anything other than a *string* data response rule, the XML schema definition (*.XSD*) which contains the response type must be manually edited.

Also, a rule project can contain multiple rules, each returning different types. It is necessary to maintain multiple versions of the *.XSD* to support this (for example, a version where the data response variable has been set to numeric, another for string, another for boolean, etc.).

The *.XSD* file downloaded from IBM Analytical Decision Management contains two complex data types: one representing the IBM Analytical Decision Management project data model which should be used for the input parameter of IBM Operational Decision Management rules projects, and one representing a rules return type which should be used for the output parameter of IBM Operational Decision Management rules projects.

The response-related *.XSD* file downloaded from IBM SPSS Collaboration and Deployment Services Deployment Manager contains the response type definition and will only contain the one complex data type.

In both of the above cases, the return type (whose name will always have a "response" suffix) defines a `value` element which defaults to a *string*. If you want to create a rules project with a different kind of return value, you must manually edit the downloaded *.XSD* before importing the IBM Operational Decision Management rules project so that the return type's `value` element is defined as desired. Following are a few examples:

```
    <xsd:complexType name="claim_dataresponse">
        <xsd:annotation>
            <xsd:appinfo>
                <dmname>claim_dataresponse</dmname>
            </xsd:appinfo>
        </xsd:annotation>
        <xsd:sequence>
            <xsd:element name="value" type="xsd:string"/>
        </xsd:sequence>
    </xsd:complexType>
```

```
<xsd:complexType name="claim_dataresponse">
    <xsd:annotation>
        <xsd:appinfo>
            <dmname>claim_dataresponse</dmname>
        </xsd:appinfo>
    </xsd:annotation>
    <xsd:sequence>
        <xsd:element name="value" type="xsd:boolean"/>
    </xsd:sequence>
</xsd:complexType>


<xsd:complexType name="claim_dataresponse">
    <xsd:annotation>
        <xsd:appinfo>
            <dmname>claim_dataresponse</dmname>
        </xsd:appinfo>
    </xsd:annotation>
    <xsd:sequence>
        <xsd:element name="value" type="xsd:long"/>
    </xsd:sequence>
</xsd:complexType>
```

Sample files

Data and other sample files are included with IBM Analytical Decision Management that can be used to replicate the examples shown in the user guides.

Data

During installation, a number of sample data files are installed under the IBM SPSS Modeler Server *\Demos* directory (for example, *C:\Program Files\IBM\SPSS\ModelerServer\17\Demos*). This data may be used with the prebuilt sample applications, and include:

- *bank_customer_data.txt*
- *bank_response_data.txt*
- *customer_contact_history.txt*
- *customer_summary_data.txt*
- *insurance_claim_data.txt*
- *insurance_fraud_data.txt*
- *inventory_orders.txt*
- *inventory_stock.txt*
- *machine_log_data.txt*
- *machine_summary_data.txt*
- *retail_purchase_data.txt*

Sample application files, models, and rules that use these data files assume the default location indicated above. If data files are installed to a different drive or location, data source paths must be updated in the sample applications accordingly.

Samples

A number of sample application files, models, and rules are provided for use with the prebuilt IBM Analytical Decision Management applications. The samples are packaged in the *\demo* directory on the installation disk, or provided in a file available for electronic download. Complete the following steps to import sample files for one or more prebuilt applications into the IBM SPSS Collaboration and Deployment Services Repository.

1. **DVD.** Copy the file *DecisionManagementExamples.pes* from the *\demo* folder on the DVD to a temporary location.
   OR

   Downloaded file. Download the file *ADM_17_Demo.zip* to a temporary location and extract it.
2. Open IBM SPSS Collaboration and Deployment Services Deployment Manager and log on to your Content Server connection.
3. Right-click the Content Repository folder and select **Import**.
4. In the Import dialog box, click **Browse** and select the *DecisionManagementExamples.pes* file you copied or extracted to a temporary location in step 1. Click **OK**.
5. In the Import Conflicts dialog, click **OK**. When the import completes, you will see the following new folders containing example projects that can be opened in IBM Analytical Decision Management.
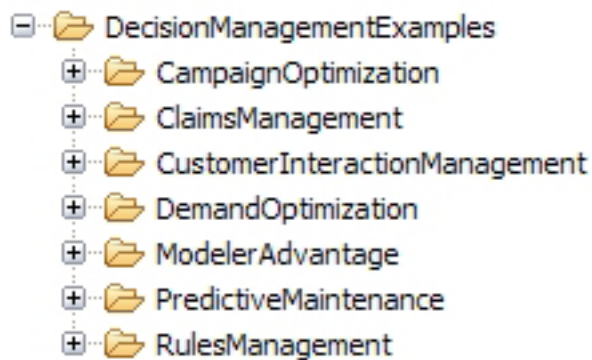
Figure 8. Folders for sample artifacts

Note that all sample application files, models, and rules assume that data files are installed under IBM SPSS Modeler Server's *Demos* directory (for example, *C:\Program Files\IBM\SPSS\ModelerServer\17\ Demos*). If data files are installed to a different location, the data source paths must be updated in the sample applications for the local configuration.

Chapter 7. Getting help

Three levels of user-assistance are provided:

- Custom help is provided for each prebuilt application packaged with IBM Analytical Decision Management. To launch the help, click the help icon on any screen or dialog:

- Coach text, which is displayed in a small pop-up window, providing details on the current application or business problem. Coach text provides an additional layer of customized, context-sensitive help that can be easily configured for a specific site or application. Coach text appears throughout the user interface when you click the small coach text icon:

- Standard tool tips, which are provided for individual controls within a page or dialog. To display a tool tip, hover the button, link, or other control.

Available documentation

IBM Analytical Decision Management documentation is available in two formats. When you access help in the application, the online IBM Knowledge Center is launched, with the relevant help topic displayed. If you have trouble launching the help, contact your administrator in case you do not have authority to access the external Knowledge Center URL.

The Knowledge Center is an online, centralized collection of all documentation. It is always available, even when the product is not running. We recommend bookmarking it in your browser. It also contains documentation for other IBM products such as IBM SPSS Modeler, IBM SPSS Collaboration and Deployment Services, and Statistics.

At times when you are without an Internet connection, or when the Knowledge Center may be temporarily unavailable for maintenance, all documentation is also available in Portable Document Format (*.pdf*). The PDF files are available on the installation media, and are also installed into the IBM SPSS Collaboration and Deployment Services installation directory (for example, *C:\Program Files\IBM\SPSS\Deployment\7.0\Server\documentation\DecisionManagement*).

PDF documentation is also available on the Web at http://www-01.ibm.com/support/docview.wss?uid=swg27043687, and release notes are available at http://www-01.ibm.com/support/docview.wss?uid=swg27043686.

Appendix. Accessibility

Accessibility features help users who have a disability, such as restricted mobility or limited vision, to use information technology products successfully. IBM strives to provide products with usable access for everyone, regardless of age or ability. This product uses standard web browser navigation keys.

Because IBM Analytical Decision Management is a browser-based application, accessibility options for disabled users, such as those who are visually impaired, are controlled from your web browser settings. For details about user interface actions, including keyboard shortcuts, screen readers, and so on, refer to your web browser's documentation.

Vendor software

IBM Analytical Decision Management may require use of some vendor software that is not covered under the IBM license agreement. IBM makes no representation about the accessibility features of these products. Contact the vendor for the accessibility information about its products.

IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment IBM has to accessibility.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who want to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Other product and service names might be trademarks of IBM or other companies.

Index