IBM® SPSS® Modeler Advantage 7

IBM.®

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 incorporating 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.

Contents

About IBM Analytical Decision Management

What's new

IBM® Analytical Decision Management 7 adds the following new 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 Analytical 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 Analytical 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 Analytical 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 ILOG Business Rule Management System integration. In IBM® SPSS® Collaboration and Deployment Services Deployment Manager, you can use the Add New External Rule dialog box to define references to ILOG rules that can then be used in IBM Analytical Decision Management. Previously, these rules could only be accessed through a web service. With this release, you can now reference local rules in the ILOG 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 Analytical 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 SPSS Modeler stream file from the local file system and store it to the repository if desired.

Multiple SPSS Modeler Server credentials. If you have multiple SPSS Modeler Server credentials defined in Deployment Manager, you can now also specify more than one for use with IBM Analytical Decision Management. Then when the IBM Analytical 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 Analytical Decision Management. This provides better control over which users have access to which parts of the SPSS Modeler Server file system.

The previous release, IBM SPSS Decision Management 6.2, added the following features:

Support for Cognos data. When used with SPSS Modeler Server 14.2 or newer, IBM Analytical Decision Management adds the ability 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. For more information, see the topic IBM Cognos BI source in Chapter 4 on p. 31.

ILOG Business Rule Management System integration. Rules created in a Business Rules Management System such as ILOG 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. These external rules can be selected and used in the same manner as other models and rules within the IBM Analytical Decision Management application. For more information, see the topic Using rules from ILOG Business Rule Management System in Chapter 5 on p. 42.

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

About IBM Analytical Decision Management

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.

Customer response modeling example

IBM SPSS Modeler Advantage

IBM® SPSS® Modeler Advantage is an easy-to-use application that puts the power of predictive modeling in the hands of business users. Using predictive models, you can identify patterns based on what has happened in the past, and use them to predict what is likely to happen in the future.

For example, you can 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. The resulting predictions can be used to generate lists of target customers or cases of interest, as input for strategic planning, or can be integrated with rules in the context of a predictive application.

Maximizing profit for direct mail campaigns

For this example of a predictive model, you are marketing financial products and services to customers at a fictitious bank. In a recent test campaign, a number of customers were offered one of several new services, and their responses recorded. However, response rates were lower than hoped and the campaign failed to turn a profit.

Using IBM® SPSS® Modeler Advantage, your goal is to turn that loss into a profit by targeting the best available customers given your expected costs, revenue, and other factors.

Results from the test campaign are summarized below.

Table 2-1
Test campaign results

Test campaign results	
Mail count	3000
Response count	111
Response rate	3.7%
Cost	6750
Revenue	6660
Total profit (loss)	(90)

This example uses the data files *bank_customer_data.txt* and *bank_response_data.txt*, which are distributed and installed with IBM® Analytical Decision Management. A completed version of this example is also provided in the file *bank_response_model.str*. Contact your administrator for details on installing and sample files if necessary. For more information, see the topic Sample files on p. 5.

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\15\Demos*). These files may be used with prebuilt sample applications, and include:

- *insurance_claim_data.txt*
- *insurance_fraud_data.txt*
- *bank_customer_data.txt*
- *bank_response_data.txt*
- *retail_purchase_data.txt*
- *customer_summary_data.txt*
- *customer_contact_history.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, the data source paths will need to be updated 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 IBM® SPSS® Collaboration and Deployment Services Repository.

► **DVD.** Copy the *DecisionManagementExamples.pes* file from the *\Demo* folder on the DVD to a temporary location.

or

► **Downloaded file.** Download the file *SPSS_Decision_Mgmt_70_demo.zip* to a temporary location and extract it.

► Open IBM® SPSS® Collaboration and Deployment Services Deployment Manager and log on to your Content Server connection.

► Right-click the Content Repository folder and select Import.

► In the Import dialog box, click Browse and select the *DecisionManagementExamples.pes* file you copied to a temporary location previously. Click OK.

▶ 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 2-1
Folders for sample artifacts
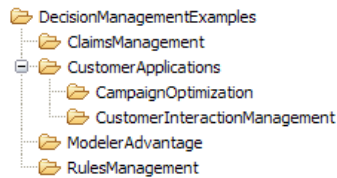
Note that all sample application files, models, and rules assume that data files are installed under the IBM® SPSS® Modeler Server   *\Demos* directory (for example, *C:\Program Files\IBM\SPSS\ModelerServer\15\Demos*). If data files are installed to a different drive or location, the data source paths will need to be updated for the local configuration.

Launching the application

The IBM® Analytical Decision Management launch page includes links to available applications at your site. These links may be configured by your administrator; you may be able choose from additional applications by clicking the Add application link in the upper right corner.

▶ To start working through the example, open the IBM Analytical Decision Management launch page. For ease of access in the future, you can add a bookmark to this page in your browser, or create a shortcut on your desktop.

▶ To create a new model, select New from the application drop-down list in the IBM® SPSS® Modeler Advantage group and click Go. The IBM SPSS Modeler Advantage Home page will be displayed.

Figure 2-2
IBM SPSS Modeler Advantage Home page

Each icon on the Home page represents a different step in the modeling process, from defining data to building and scoring the model. To get started, click the Data icon at left.

Setting up data sources

The Data tab allows you to configure data sources for the sample data files used in the example. Data sets can be configured by your administrator, or you can add your own.

The project data model defines the fields that are available throughout the application, and may be required for rules and models to evaluate. All other data sources must be mapped to the data model in order to be used with the application.

This example uses two data files, both of which are distributed with the application:

- **bank_customer_data.txt** includes information about current customers, including customer ID, payment history, the number of products they own, the number of months they have been a customer, as well as demographic information like age and income. The project data model is based on this file.

- **bank_response_data.txt** contains information about customers who have been offered one or more services in previous campaigns, and records their responses. By analyzing this data, the application can identify which customers are most likely to respond to similar offers in future campaigns.

By default, the sample data files are installed in the IBM® SPSS® Modeler Server *Demos* folder accessible from your application (for example, *C:\Program Files\IBM\SPSS\ModelerServer\15\Demos*. Contact your administrator for assistance installing these files if necessary.

To define the project data model

▶ On the Project Data Sources panel, click Add a data source.

▶ Specify *bank customer data* as the data source name.

▶ Select File, and select *bank_customer_data.txt*.

▶ Under Specify input fields, make sure the measurement type for the *Has Bad Payment Record* field is set to Flag. The flag setting indicates that the field can have two possible values, indicating *true* and *false* (typically coded as *T* and *F*, or 0 and 1 in a numeric field), to indicate the presence or absence of that attribute.

▶ Under Specify input fields, also make sure the measurement type for the *Gender* field is set to Nominal. (The values of this field simply indicate gender as male or female, and not *true/false* as would be the case with a flag.)

▶ Click Save.

Since this is the first data source defined for the project, it is automatically selected for the project data model, and used to define the fields required by the application. All other data sets are mapped relative to this source.

To define the response data source

The response data set contains many of the same fields used in the project data model, as well as two additional fields named *Campaign ID* and *Response*. These fields indicate which offer was made to each customer, and whether or not they responded, and can be used to build a model that predicts which customers are most likely to respond in the future.

▶ On the Project Data Sources panel, select Add a data source.

▶ Specify *bank response data* as the data source name.

▶ Select File, and select *bank_response_data.txt*.

▶ Under Specify input fields, make sure the measurement type for the *Response* field is set to Flag.

▶ Under Specify input fields, deselect the check box next to the *Campaign ID* field to indicate this field should not be used as an input when modeling. (This field can be used to indicate which campaigns a customer was selected for in the past, but is not useful in predicting future response.)

▶ Click Save.

Because some of the fields in the data model are not found in the response data set, an orange icon is displayed for this data source under the Compatible column on the Data tab, indicating that some fields match, but not all. The data set is suitable for modeling, but not for simulation, scoring, or deployment. (To be used for these purposes, all fields from the data model need to be present.)

Building the model

The Modeling tab allows you to specify the data source, target field, and other options to build and evaluate the model.

▶ Select *bank response data* for the data source, and select *Response* as the target field.

▶ If a Data scan required message appears, click the icon to perform a data scan. Then click Build Model.

When the model completes, a distribution chart is displayed. The chart shows observed values on the horizontal axis, with a value of 1 indicating a positive response.

■ The height of the bars reflects the overall response rate. Out of 3000 customers in the response data set, 111 responded, for an overall hit rate of about 3.8%.

■ The colored areas within each bar represent the responses predicted by the model, with red indicating a positive prediction, and blue indicating a negative response, as shown in the legend in the upper right corner. Due to the low overall response rate, there are relatively few predicted responses (red areas). However, the false bar (*Response = 0*) is overwhelmingly blue, indicating that the overwhelming majority of those predicted not to respond actually did not.

▶ To get a better idea of how the model will perform, proceed to the next section (Evaluating the model).

Evaluating the model

From the Modeling tab, click Evaluate to see how the model performs when applied to your customer data, and to understand how using the model can impact your bottom line.

▶ For the data source, select *bank response data*.

▶ Select Gains chart and Profit chart for the graph types to build.

▶ Select Simulate maximum profit, and specify *3000* for the population. This matches the number from your test campaign.

▶ Specify *2.25* for cost and *60* for revenue. These values indicate the cost to you for each offer made (such as production, handling, and mailing costs), as well as the expected revenue from each favorable response.

▶ For Response sought, select *1*.

▶ Click Run.

The gains chart shows the improvement or "lift" that can be obtained by using the model. Because the Automatically partition data option was selected (under Optional settings) when building the model, separate charts are displayed for the training and testing partitions. Each shows the proportion of hits (positive responses) at each percentile relative to the total number of hits in the sample. The flat diagonal line plots the expected response if no model were used—to gather half the hits, you would need to contact half the people. The curved line shows the improvement using the model—by targeting the 60% of customers who are most likely to respond, you can gain almost all the hits with a much smaller mailing.

The profit chart shows the gains in terms of overall profit, based on the cost and revenue values specified. For the sample used to train the model, the profit appears to top out just under the 30th percentile.

To understand how this works, go to the Simulate tab to see the expected maximum profit for the size of the population you specified. Expected profit is computed using the following formula:

```
expected profit = (probability to respond * revenue) - cost
```

Where the `probability to respond` is determined for each customer based on the model. By targeting those identified by the model as most likely to respond, you can maximize your profit per customer. As you include more customers, the total profit will increase up to a certain point, even as the return per customer gradually decreases as customers with lower response rates are included. The simulation helps you decide where to draw the line to obtain the maximum profit.

The profit comparison summarizes this result in more detail.

■ **Targeted 28%.** This column shows the result from targeting the 28% of customers with the highest rates as predicted by the model. By selectively including the 840 customers most likely respond, you can maximize your return, resulting in an expected profit of $3,450.00.

■ **Random 28%.** Shows the result if you drew a random sample of the same size. In this case you would send the same number of offers (840), but the expected response would be lower (matching the rate from the test campaign) because you are not benefitting from the extra information provided by the model. This results in a loss, because the overall response rate is not high enough to return the cost of mailing each offer.

■ **Everyone.** Shows the result if you included all 3000 customers in the sample, exactly matching the result from your test campaign. Again, this returns a loss because you are not benefitting from the model.

Looking at the results, you can see how using the model can improve the response rate to turn a losing campaign into a profitable one. The difference between a $90.00 loss and a profit of $3,450.00 represents the benefit gained through predictive modeling. This return is based on the 3000 customers from your test campaign, but by increasing the size of your target population, the benefits increase accordingly.

▶ After reviewing the simulation results, select Use for Scoring under the Predicted Profit chart to update the Score tab settings based on the specified percentage (28% in this case). For more information, see the topic Scoring the model on p. 11.

Predictive rule modeling

While predictive models may deliver all the predictive power you need, there are times when you may want to gain a little more control and insight into the inner workings of the model. In this case predictive rule (interactive) modeling can be used. With predictive rule models, you can manually create rules to segment your data, or automatically find segments with high or low concentrations of the value of your chosen field. For example, you might look for customers who gave a positive response to your campaign and then identify segments with a higher probability of responding.

▶ Returning to the Modeling tab, click Change Model and select Predictive rule model.

▶ Select *bank response data* for the data source, and select *Response* as the target field.

▶ Specify 1 as the response sought, indicating you want to look for customers who gave a positive response.

▶ Select Start Build.

▶ Select Grow Model to identify segments with a higher probability of responding. The three best rules that identify segments with the highest probability to respond are created and displayed.

▶ Click the Include/Exclude icons to select whether or not to include or exclude customers selected by each rule. For example, the second and third rule segments show response rates just over 8%, which is still higher than the overall test campaign rate of 3.7%, but not nearly as high as the rate of 15.65% for the first segment. To see how including or excluding these two rules may impact your overall profit, you can run the profit simulation both ways and compare the result, as follows:

▶ With all three rule segments set to Include, click Evaluate.

▶ Select Gains chart and Profit chart for the graph types to build.

▶ Select Simulate maximum profit and specify 3000 for the population, 2.25 for cost, and 60 for revenue. These are the same specifications used when evaluating the predictive model. The response sought is automatically completed.

▶ Click Run and, when the graphs are built, go to the Simulate tab.

The maximum simulated profit for this model is $3172.00, a bit lower than for the automated model in this case. Optionally, you can close the Evaluate dialog, change the setting for the second and third rule segments to Exclude, and repeat the profit simulation to see how this changes

the result. The real benefit of predictive rule (interactive) modeling is the increased control and insight it affords.

Testing the model

When you have built your model, you can test a selection of customer records to see how they are likely to respond to your campaign.

▶ To preview how specific records will be processed by the model scoring, select Test from the Modeling tab. The Test dialog is displayed.

▶ In the Records area, increase the Return figure to 100.

▶ In the Select Fields to Display area, select *Customer ID*. In addition, make sure the *$XF-Response* and *$XFRP-Response* fields are selected, as these indicate the predicted response (0 for no, 1 for yes in this case) and propensity score for each customer.

Optionally, you can include additional fields such as *Age*, *Weeks Since Last Offer*, and *Has Bad Payment Record* that may be useful in understand how or why a given customer will be scored a certain way by the model.

Scoring the model

To score a model means to apply it to the data source that you are interested in. For example, having generated a model that predicts which customers are most likely to respond, you can now score prospective customers to identify who you want to include in your campaign. In the Evaluate dialog box, you determined the percentage of records to score in order to maximize profit, and indicated the percentage of records you wanted to score based on this result.

▶ Select the Score tab.

▶ Under Specify output fields, ensure the *Customer ID* field is selected.

▶ Specify a destination for the scoring results. This can be either a database table or file. (If you just want to preview the scores, you can skip this step.)

▶ Under Scoring, the settings should already be configured to score the top 28% of records, based on the profit simulation you ran in the while evaluating the model. (If you did not select the Use for Scoring button when viewing the simulation results, you can update the settings here.)

▶ At the bottom of the dialog, click Preview to display a sample of the results your scoring will obtain.

Because you have chosen to score based on propensity, only customers with the highest propensity to respond are included. By linking the customer ID back to your customer database, you could generate a mailing list based on this result.

▶ Click Score; the model is applied to the customer data and a message is displayed at the bottom of the dialog to confirm this.

Cluster modeling example

With clustering models, you can categorize records into groups with similar characteristics. This can help find natural groups in your data. For example, you might segment customers based on demographic characteristics or purchasing behavior.

This example uses the data file *bank_customer_data.txt*, which is distributed and installed with the product. A completed version of this example is also provided in the file *bank_cluster_model.str*. Contact your administrator for details on installing and sample files if necessary. For more information, see the topic Sample files on p. 5.

▶ Return to the application Launch page and create a new IBM® SPSS® Modeler Advantage project.

▶ On the Data tab, in the Project Data Sources panel, add a new data source called *bank customer data* as described in Setting up data sources on p. 6.

▶ On the Modeling tab, click Change Model, select Clustering model, and click Save.

▶ For the data source, select *bank customer data*.

▶ Expand the Optional Settings section. For clustering models, additional Auto Cluster Options optional settings are available for selecting an evaluation field or setting a desired range of clusters to find.

When a clustering model is built, behind the scenes a number of different clustering models are actually created. These models are then compared and ranked by some measure of quality, and the best model is selected for use by the auto cluster model. So these two Auto Cluster Options settings provide some control over which of the models is actually selected. If an evaluation field is specified here, the model which best differentiates values of the evaluation field will be selected. And if the desired range is set, any models which find a number of clusters outside the specified range will be discarded.

▶ For this example, select the Set a desired range for the number of clusters found option and use the default minimum of 3 and maximum of 15. This will ensure the clustering model won't have too few or too many clusters.

▶ The rest of the optional settings are same as for other modeling types. Deselect the Automatically clean up and prepare data for reliable model building option just to make the auto clustering example easier to understand.

▶ Collapse the Optional Settings panel and expand the Clusters panel.

In the Clusters panel, there's a section for manual clusters and a section for auto clusters. Manual clusters allow you to define clusters based on your knowledge of the data. For this example, we'll define a manual cluster for *high value customers* and a manual cluster for *young single customers*, as follows.

▶ Click the Create a new rule icon. Name the rule *high value customers*, add the following to expressions, and click OK.

```
Months as a Customer > 12

Income > 65000
```

▶ Create another new rule called *young single customers* with the following expressions.

```
Age < 35

Marital Status = U
```

▶ The two new rules will be listed under the Manual Clusters section. To see how many customers are being caught by these manual clusters, click Clusters record count. Then expand the manual clusters again and look at the Count column to see how many customers were found for each cluster. You'll see that 67 high value customers were found and 33 young single customers were found.

▶ Next click Find Auto Clusters to automatically find any other clusters that might be of interest. When finished, expand the Auto Clusters section. You can see that three auto clusters were found.

▶ Click View Auto Clusters to see more detail. The Auto Cluster Results Viewer will be displayed.

The *Model Summary* page provides basic information about the auto cluster model algorithm that was used (*TwoStep*, in this example), the number of auto clusters found, and a basic idea of cluster quality.

Other charts are also available. You can hover over various areas in each chart to see more detail, and interact with some of the charts (such as the *Predictor Importance* chart).

Using the *Clusters* view (the last chart), you can also compare clusters to look for interesting patterns. In a real world implementation, you would study the auto clusters and then give them more meaningful names (by clicking each cluster label in the *Clusters* view). After updating the cluster label names, the new names will be used in all other charts and back on the Modeling tab. Then you would proceed to evaluate and test the model, and finally score it once you're happy with it.

For more details about cluster modeling, see Building a clustering model on p. 47.

Association modeling example

Association modeling is an important technique in predictive analytics. With association models, you can predict which items are most likely to appear together, as well as the strength of the relationships between them. For example, you might identify the next likely purchase for a customer based on items in their current basket.

This example uses the data file *retail_purchase_data.txt*, which is distributed and installed with the product. The data consists of purchase information for electronics items such as televisions, computers, and smart phones.

A completed version of this example is also provided in the file *retail_association_model.str*. Contact your administrator for details on installing the sample files if necessary. For more information, see the topic Sample files on p. 5.

▶ Return to the application Launch page and create a new IBM® SPSS® Modeler Advantage project.

▶ On the Data tab, in the Project Data Sources panel, add a new data source called *retail purchase data*. Select File and select *retail_purchase_data.txt*.

▶ On the Modeling tab, click Change Model, select Association model, and click Save.

▶ For the data source, select *retail purchase data*.

▶ For the data format, select Tabular. A transactional data set would only show items that were included, and all items would appear in a single column. But in this data set, each item is in a separate column with a *true* or *false* flag to indicate whether it was purchased or not.

▶ Click Build Model. You will receive an error that says building failed and suggests that the thresholds may be too high. Close the error.

When looking for association rules, it's usually a good idea to look for rules that are generally applicable, reliable, and simple.

▶ Expand the Optional Settings section. Under Association Rule Options, change the first threshold from *10* to *1*, and lower the second threshold from *80* to *1*. Normally, you would experiment with different thresholds, rebuilding the model each time, to strike a balance between not getting too many rules and not getting enough.

▶ Click Build Model.

In the model results, each row represents a separate association rule. The first rule in the list says that if a customer purchases a *Smart Phone*, they're also likely to purchase a *Big Screen TV*. The Support column shows that about 15% (or about 1 in 6 customers) has purchased the *Smart Phone*, and the Confidence column shows that of those customers, there's about a 30% probability that they'll also purchase a *Big Screen TV*.

So even with a small set of rules (239 in this example), you might be able to use them in a business context to predict or even recommend what customers will buy next.

▶ Next, click Test and then click Run to see how the predictions or recommendations will be made.

In this case, the model looked at the rules for the highest confidence figures, looked at what the user has already purchased, and then on the basis of these two together, has made some recommendations about what customers are likely to purchase.

▶ Click the preview icon beside the first record, for example, to see that the first customer purchased a *Big Screen TV* and *Speakers*, but has not purchased any other products. Close the preview.

The recommendation for this customer is to offer them a *Standard TV*, as seen in the second column. The third column's value. .29, means there's a 29% chance that customers who have already purchased a *Big Screen TV* and *Speakers* will also purchase a *Standard TV*.

However, how do we know whether this is an anomaly or a predicted trend that can be generally applicable? Take note of the rule number (115 in this example) so we can go back and look at the rule in more detail.

▶ Close the Test dialog.

▶ Page through the model results until you see rule number *115*. Note that for large models with many more pages of results, you can also click Find a rule by ID and type the rule number.

▶ Click the arrow next to any one of the result column headers and select
Columns > Instances

The Instances column will be displayed. Based on the number of instances, it is clear this is not an anomaly because rule 115 is based on 58 different customers who have purchased the three products together. This is enough to give us confidence about using the rule as the basis for future recommendations.

We have looked at a single rule to see how it might be used for a recommendation. But IBM® Analytical Decision Management also provides powerful options for customizing the model to make it more relevant to our specific purposes. For example, we may want to predict which customers are interested in buying a *Big Screen TV*:

► Click Apply Filters. This displays a dialog box where you can define filters so that only rules matching the filters will be shown in the model.

► Select Enable filter and select *Big Screen TV* for the item.

► Click Save and click Save again to return to the model results.

The number of rules has dropped from 239 to 34.

You might also want to confirm that you're only using rules that are based on a reasonable number of customers. To achieve this, you might exclude rules that are based on the behavior of less than 10 customers:

► Click the arrow next to the Instances column header and select Sort Ascending.

► In the Exclude column, select all rules with under 10 instances.

► Note that even though many of the rules have now been filtered and excluded, they're still in the model at this point. To remove them permanently so they're not used when the model is scored, click Create New Model - deleted excluded & filtered rules and click OK.

The number of rules returned is reduced to 20, with all of them based on 10 or more instances, and all predicting *Big Screen TV*.

► To see how many of these rules will be applied, click Test and then click Run. *Big Screen TV* is now being recommended for several more customers, and some have a higher confidence level than others.

► Close the Test dialog box.

► In the Optional Settings section, expand the Scoring Options section to display scoring options specific to association models.

By default, the maximum number of predictions is 3. This will be based on the highest confidence figures, or you can change the rule criterion to choose rules based on the widest number of instances, or choose rules based on the highest amount of *lift* (the increase in probability of a particular item being purchased). You might also want to make sure not to offer a *Big Screen TV* to customers who already purchased one by selecting Ensure predictions no present.

The default scoring options are suitable for most situations. But in some cases, the added flexibility these options provide may be advantageous.

The next step in the process would be to use the Scoring tab to score the model. For more information, see the topic Scoring models to a database table, file, or Cognos BI server in Chapter 7 on p. 62.

For more information about association modeling, see Building an association model on p. 48.

Navigating the workspace

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.

▶ 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. For more information, see the topic Gallery on p. 19.

Project configuration

When launching an application that includes a Define tab with a dimension tree, if the selected application is configured to include more than one dimension hierarchy 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. For more information about the Define tab and its dimension tree, see .

▶ The Project configuration dialog displays all available dimension hierarchies defined in the application template. Select one or more dimension hierarchies (elements) you wish to use to define the project. A description of each element may be available as defined by the application designer in the application template.

The element name is the name of the top-most dimension. Dimensions can only belong to one element.

▶ If you select more than one element, you can sort the order in which they'll appear throughout the application and which elements can depend on each other.

The order you specify here is important. The elements will appear in ordered tabs, and hierarchy members of previous tabs can be referred to in rules within subsequent tabs, but references cannot be made to subsequent tabs. For example, rules on a second and third tab can refer to hierarchy members in the first tab, but rules on the first tab cannot refer to members on the second or third tabs (and rules on the second tab cannot refer to members on the third tab).

▶ When finished in the Project configuration dialog, click OK.

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.

Model. Build models using analytical data, evaluate them using predictor importance, distribution, gains charts, and so on to gain insight by discovering patterns hidden in data, and use these to further refine your decisions.

Score. Score the model.

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 and Simulations tabs 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 3-1
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.

Model settings

Model type build default. Specify whether by default you want the application to build the best model automatically from the available data (predictive model), if you want to use a model that identifies segments that show a higher or lower likelihood relative to a specific outcome (predictive rule model), if you want to use a model that categorizes records into a certain number of clusters (clustering model), or if you want to use a model that predicts which items are most likely to appear together and predicts the strength of the relationship between them (association model). Only model types configured to appear in the application will be available in this drop-down menu.

Automatically clean up and prepare data for reliable model building. Selecting this option enables you to let the application do a lot of the data preparation. For example, repairing data issues, screening less useful data, or deriving new attributes.

You can choose whether the clean up operation should give an equal balance to both speed and accuracy, give priority to accuracy over speed, or give priority to speed such as when you are working with very large datasets or are looking for a quick result.

Restrict time spent building a single model. Use this option to specify the default amount of time to spend building each model. This can be especially useful when processing large datasets, and can have a significant impact on the time needed for model building to complete.

Test and simulation settings

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

Opening and saving projects, models, and rules

You can open from or save your work to the IBM® SPSS® Collaboration and Deployment Services Repository, or download to the local file system. You can also upload IBM® SPSS® Modeler stream files from the local file system and save them 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 model.
	Saves the current project, model, or rule 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, model, or rule, 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, model, or rule from the IBM SPSS Collaboration and Deployment Services Repository and/or local file system as applicable.
	Uploads a project, model, or rule from the local file system. File system access is not supported for all applications. If uploading an SPSS Modeler stream, you can choose to open the file or save it to the repository.

Saving to the IBM SPSS Collaboration and Deployment Services 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.

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.

When , a Save 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.

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.For more information, see the topic Exporting and reusing rules in Chapter 5 on p. 40.

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. For more information, see the topic Opening and saving projects, models, and rules on p. 21.

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 for editing, the object is locked in the IBM® SPSS® Collaboration and Deployment Services Repository so that other users cannot edit it. This is indicated by a check box in the upper right corner of the workspace. Alternatively, you can deselect this check box to allow others to edit the object.

- 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 browse the workspace, but you can no longer save changes.

Object properties

When editing a ruleset, model, or project, you can add keywords and a description to help identify the object when browsing for it at a later date. To display the Edit Properties dialog, click the Properties icon in any application.

Figure 3-2
Properties icon

Keywords. Enter any keywords to quickly identify the saved object when searching for it. For example these might be "Sales" and "Q1-2012."

Description. Enter a longer description of the object saved. For example, this might be "Analysis of retail sales; first quarter of 2012."

Managing data sources

The Data tab enables you 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 tab is divided into three 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 will use. If desired, the Additional Fields section allows you to add expressions or model output from different data sources to extend your project data model.

- **Project Data Sources.** Lists the data sources that have been saved for use with the current project, model, or rule.

- **My Data Sources.** Lists data sources you have defined or saved. Data sets from this list can be copied to or from different projects, 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.

Key. If the data sources selected for the application is an Application View data source and Application View keys have been defined on the data source table, then this drop-down will provide a list of keys to choose from. Application View keys are part of the IBM® SPSS® Collaboration and Deployment Services Enterprise View created in IBM® SPSS® Collaboration and Deployment Services Deployment Manager. If there are no keys defined for the Application View table, or if the data source selected for the application is not an Application View data source, then a list of all operational data fields from the data source will be available to choose from in the Key drop-down instead.

The selected key field will be used to map to any additional data source tables defined in the Additional Fields section. At run time, the keys are used to join the primary Project Data Model table with any secondary tables. The key field in the Project Data Model is assumed to be unique (for example, a customer ID, claim ID, etc.). Each secondary table must have a matching column.

The key field is optional if there are no additional tables defined for data source.

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. For more information, see the topic Measurement levels on p. 27.

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.

Add/Edit additional fields. This section allows you to add expressions or model output from different data sources to extend your project data model, if desired, by clicking Add/Edit additional fields.

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 will be shown beside the icon.

Preview. Enables you to preview a sample of the data contained in the source. For more information, see the topic Previewing data on p. 27.

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 with the project data model the data source is.

■   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 the topic Mapping fields on p. 35.

Copy. Enables you to copy 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 has given 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 the topic Creating a new data source on p. 28.

▶ 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 the topic Measurement levels on p. 27.

▶ 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 have been mapped, the link is no longer displayed.) For more information, see the topic Mapping fields on p. 35.

▶ To preview a data source, click the Preview icon. For more information, see the topic Previewing data on p. 27.

▶ To add additional fields (expressions or model output from different data sources) to extend your project data model, click Add/Edit additional fields. For more information, see the topic Adding additional fields on p. 34.

Previewing data

To verify that you have selected the correct data source, click the Preview icon to display a sample of the data contained in the source. The preview includes any additional fields listed in the Additional Fields section of the Data tab.

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 the topic Selecting input fields on p. 34.

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*.

- **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.

- **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).

Creating a new data source

► To access the Data Source Editor, select Add a data source under either the Project Data Sources or My Data Sources area.

► Type a name for the data source.

► Select the source type:

- **File.** Many file types are supported, such as text (*.*txt*, *.*csv*, etc.), Excel, the *.*sav* format used by IBM® SPSS® Statistics, and SAS files. For more information, see the topic File source on p. 28.

- **Database.** Any database that supports ODBC, including Microsoft SQL Server, DB2, Oracle, and others. For more information, see the topic Database source on p. 31.

- **Enterprise View.** Creates a connection to an IBM® SPSS® Collaboration and Deployment Services Enterprise View as defined in IBM® SPSS® Collaboration and Deployment Services. For more information, see the topic Enterprise View source on p. 31.

- **IBM Cognos BI server.** Any database that supports IBM Cognos BI server. For more information, see the topic IBM Cognos BI source on p. 31.

► 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 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.

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.

Specify input fields. For more information, see the topic

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. For more information, see the topic Selecting input fields on p. 34.

IBM SPSS Statistics data files

When you select a IBM® SPSS® Statistics data source, you are prompted to enter further details:

Variable names. Select a method of handling variable names and labels upon import from a SPSS Statistics *.sav* 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 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 SPSS Statistics *.sav* 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* 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 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. For more information, see the topic Selecting input fields on p. 34.

Database source

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

Test connection. Enter your database username and password and click to check that you have access to the database.

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 Never if you *never* want table and field names quoted.
- Select Always if you want *all* table and field names quoted.

Specify input fields. For more information, see the topic Selecting input fields on p. 34.

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, Simulate, WhatIf? analysis, 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. For more information, see the topic Selecting input fields on p. 34.

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.

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.

▶ 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.

▶ 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.

▶ 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.

▶ 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.

▶ 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.

▶ 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.

▶ 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.

▶ 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 4-1
Cognos object icons

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

Adding additional fields

In most IBM® Analytical Decision Management applications, you can add additional fields to your project data model such as expressions or model output. The Additional Fields section of the Data tab, under Project Data Model, displays any expressions or model outputs that are to be included as part of the data. Note that the Additional Fields section is not included in the IBM® SPSS® Rules Management application.

▶ In the Project Data Model section, on the right side of the screen, click Add/Edit additional fields.

To add an expression

▶ Click Add an expression. The expression editor will be displayed. For more information, see the topic Expression editor on p. 35.

To add model output

▶ Click Add model output then browse for and select a model. The Inputs and Outputs dialog for the selected model will be displayed.

▶ If the model input fields are compatible, the Model Input Mapping section of the dialog will be collapsed. If the model input fields are not compatible, the section will be expanded and you must map available table inputs to inputs of the selected model.

▶ In the Model Output section, all available model outputs for the selected model are listed. Select the model output fields you want to add to your data. The Model output name column displays the name of the output field as generated by the model.

Any fields the model outputs can be selected. The model output fields are all the fields coming out of the stream which are not also inputs to the stream (all fields added by the stream). The input fields may not include everything in the project data model, so it's possible for a model output field to have the same name as a project data model field. So you can also change the data field names here if desired.

The Measurement column displays the measurement type of the model output prefixed with the appropriate type icon, and the Values column displays the values of the model output. Note that these columns may display results for some of the fields and not others. Typically the measurement types and values are known for actual model outputs (predicted value, probability, etc.) without a data scan, so if known it will be displayed. But the measurement types or values may not be known for other output fields from a stream (for example, a derived field) without a data scan.

▶ A Model Parameters section will be included at the bottom of the dialog if the selected stream includes scoring parameters. Enter values for the model parameters and click Save when finished in the dialog. Any selected model outputs will be added to the Additional Fields section of the Data tab.

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.

▶ 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 that you have defined are also displayed. For more information, see the topic Expression editor on p. 35.

▶ Select each of the fields and expressions that 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 the topic Measurement levels on p. 27.

Expression editor

You can use the expression manager 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 always numeric with a measurement of Continuous; this can not be changed.

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

▶ To display existing expressions, on the Data tab, click Add/Edit additional fields and then click Add an expression.

Creating an expression

▶ In the Expression editor dialog, enter a unique name for the expression.

▶ All available inputs from the data source are shown. Either double-click or drag the required inputs and mathematical symbols to the Expression Equation field to build your expression.

▶ Enter a Description that explains what the expression is used for.

▶ Click Save.

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.

▶ 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.

▶ 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.

▶ When finished, click OK.

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.

Creating rules

► 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. For more information, see the topic Exporting and reusing rules on p. 40.

► 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. For more information, see the topic Project configuration in Chapter 3 on p. 18.

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

► In the application where you want to add the rule, click the Create a new rule icon.

► Specify a name for the rule.

► For the input column, click the down-arrow icon to specify a field, rule, or model, as desired.

Table 5-1
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.

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

Table 5-2
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

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

▶ 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. For more information, see the topic

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. For more information, see the topic

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.

▶ 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.

▶ Specify a name for the rule, select whether to use the rule for inclusion or exclusion (or do this later), and add one or more expressions as desired. For more information, see the topic

▶ Click OK to save the rule.

▶ Click the icons in the Include/Exclude column if you want to toggle this setting for each rule. For referenced selection rulesets, this setting cannot be changed. You cannot combine inclusions and exclusions in the same rule.

Alternatively, click the Find an existing rule icon to browse for rules that have already been defined. For more information, see the topic

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.

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).

► 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.

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 ILOG Business Rule Management System

Rules created in a Business Rules Management System such as ILOG can 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.

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

The Modeling tab allows you to specify a data source, target field, and other options. You can also choose which model type to build.

Note that if **parameters** were added to the application (stream file) in IBM® SPSS® Modeler, then a Parameters link will be available on the Modeling tab and the Score tab. Clicking the Parameters link shows the defined parameters and their current values, and allows you to change the parameter values to affect the model build or score.

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

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.

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

▶ 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.

▶ Specify optional settings as desired. For more information, see the topic Optional model settings on p. 50.

▶ 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. For more information, see the topic Gallery in Chapter 3 on p. 19. 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, distribution and predictor importance charts are displayed. For more information, see the topic Predictive modeling results on p. 44.

▶ Optionally, use the Evaluate and Test features to see how the model performs on your sample data. For more information, see the topic Evaluating models on p. 53.

Predictive modeling results

When a predictive model completes, distribution and variable importance charts are displayed, providing a quick visual assessment of model performance, and of the importance of different predictors in obtaining the final result.

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.

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

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

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

▶ 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.

▶ Specify optional settings as desired. For more information, see the topic Optional model settings on p. 50.

▶ Click Start Build.

▶ 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.

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

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

▶ 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.

▶ Optionally, you can manually add or edit segments as desired. For more information, see the topic Defining segment rules in Chapter 5 on p. 37.

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

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

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

▶ Specify optional settings as desired. For more information, see the topic Optional model settings on p. 50.

▶ Click Find Clusters.

▶ Optionally, you can add manual clusters. For more information, see the topic Using manual clusters on p. 48.

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

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.

▶ In the Manual Cluster section, click the Create a new rule icon and add one or more rules to define manual clusters.

▶ If desired, change cluster label names.

▶ 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

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

▶ Specify a data format. For more information, see the topic Tabular versus transactional data on p. 49.

▶ Specify optional settings as desired. For more information, see the topic Optional model settings on p. 50.

▶ Click Build Model.

▶ 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.

Tabular versus transactional data

When building an association model, Tabular or Transactional data formats are available. Tabular data is selected by default. If you choose Transactional, select fields for ID and Content from the drop-down menus that appear.

Transactional Format

Transactional data have a separate record for each transaction or item. If a customer makes multiple purchases, for example, each would be a separate record, with associated items linked by a customer ID. This is also sometimes known as **till-roll** format.

Customer	Purchase
1	jam
2	milk
3	jam
3	bread
4	jam

Customer	Purchase
4	bread
4	milk

Tabular Data

Tabular data (also known as **basket** or **truth-table** data) have items represented by separate flags, where each flag field represents the presence or absence of a specific item. Each record represents a complete set of associated items. Flag fields can be categorical or numeric, although certain models may have more specific requirements.

Customer	Jam	Bread	Milk
1	T	F	F
2	F	F	T
3	T	T	F
4	T	T	T

Optional model settings

When building models, you can choose from the following optional settings. If changing any of these settings for an existing model, the model must be rebuilt for changes to apply.

Automatically partition data, to enable model validation on build data source. 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.

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.

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. For more information, see the topic Expression editor in Chapter 4 on p. 35.

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. For more information, see the topic Defining selection rules in Chapter 5 on p. 39.

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 distribution, gains, lift, response, profit, and ROI charts, and to simulate the maximum profit possible. This dialog is only available if the selected model is complete. Note that evaluation is not available for the Association model type.

Figure 6-1
Gains chart
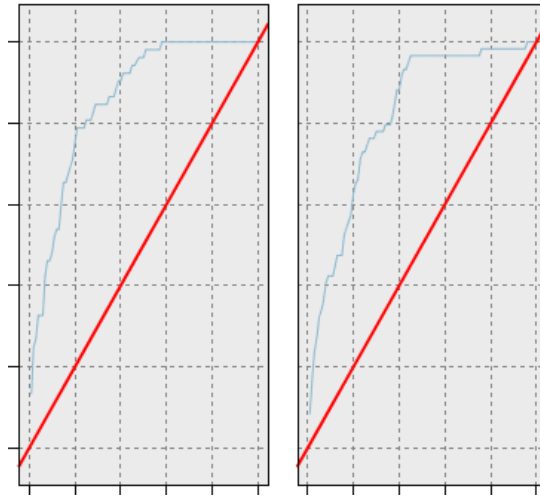
▶ In the upper panel, select an evaluation data set. As with model building, this must be an analytical data set for which actual responses are known. (Otherwise you have no way to compare actual responses to those predicted by the model.)

▶ Select the graphs you want to build.

▶ If simulating profit, select the Simulate maximum profit box and enter the population on which to base the profit calculation. For more information, see the topic Simulating maximum profit on p. 57.

▶ 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.

▶ 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.

▶ 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.

For flag and set targets, you can choose from the following additional charts:

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.

Figure 6-2
Gains chart
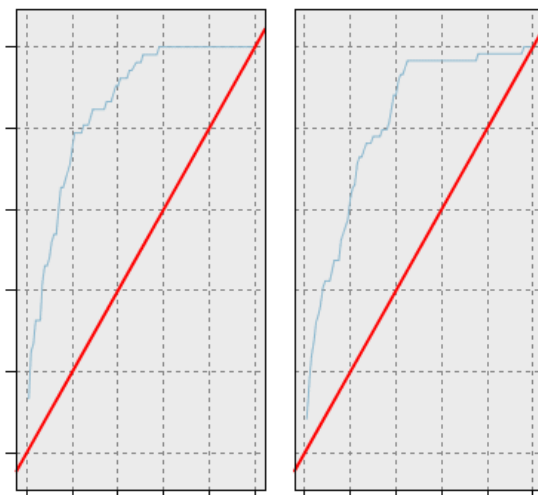
Lift charts

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.

Figure 6-3
Lift chart
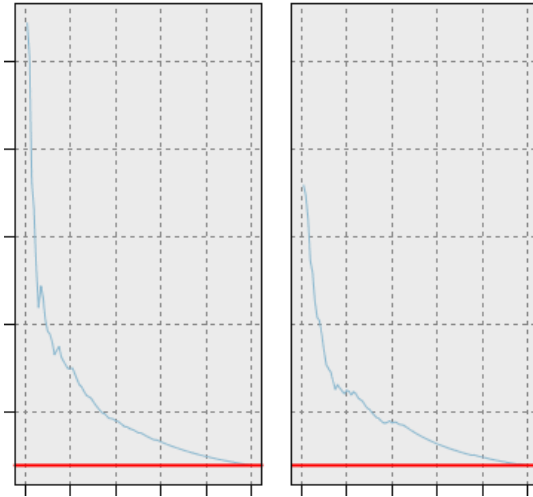
Response charts

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

Figure 6-4
Response chart
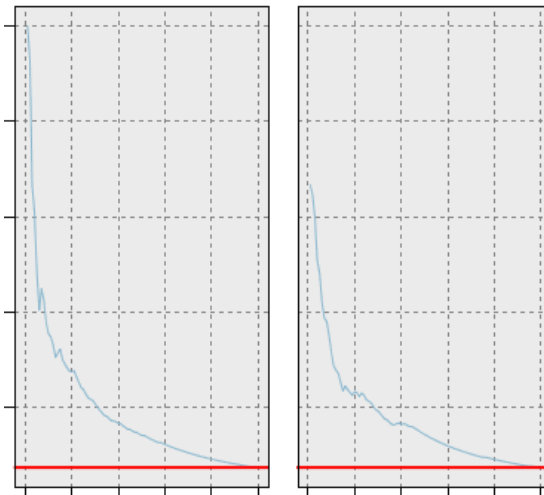
Profit charts

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. For more information, see the topic Simulating maximum profit on p. 57.

Figure 6-5
Profit chart
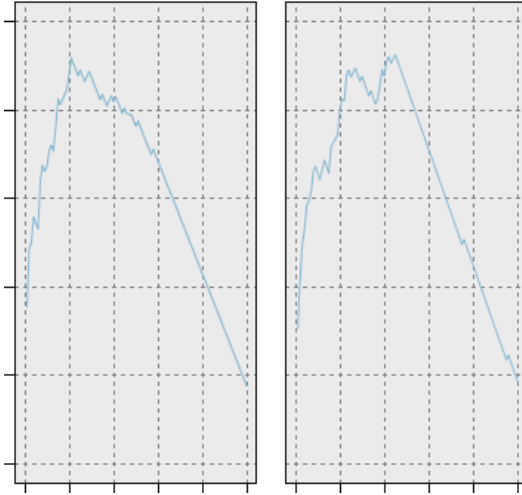
ROI charts

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

Figure 6-6
ROI chart
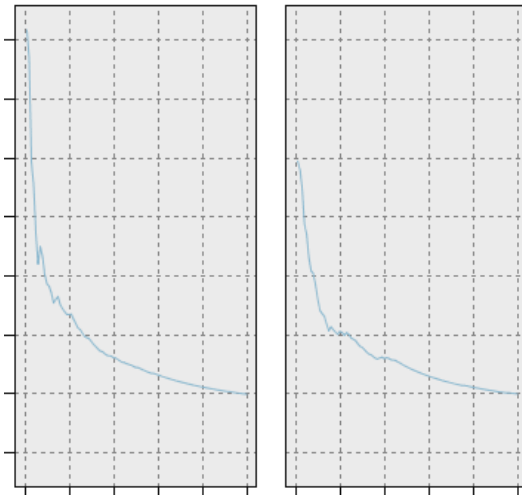
Simulating maximum profit

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.

Figure 6-7
Simulation settings
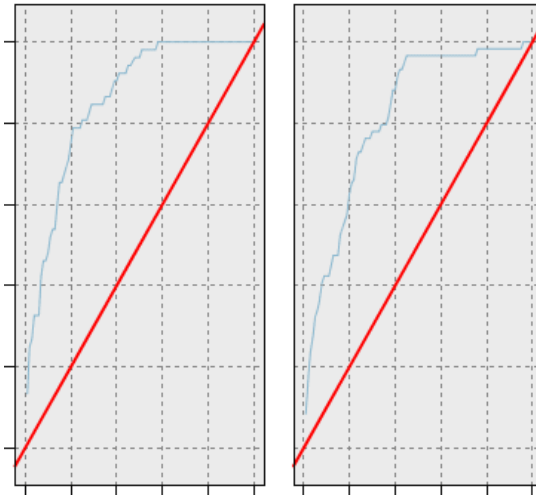
To predict the maximum profit:

▶ 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.

▶ 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.

▶ Specify the expected revenue that a positive response would generate. For example, the retail cost of an item in a campaign.

▶ Select the response you are interested in.

▶ 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. For more information, see the topic Evaluating the model in Chapter 2 on p. 8.

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.

When the figures are acceptable, you can choose to use the settings for scoring your records. To do this, click Use for Scoring, the percentage of records to target is displayed in the Top % field in the Propensity Options area of the Score tab. For more information, see the topic Scoring models to a database table, file, or Cognos BI server in Chapter 7 on p. 62.

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.

Figure 6-8
Cumulative profit chart
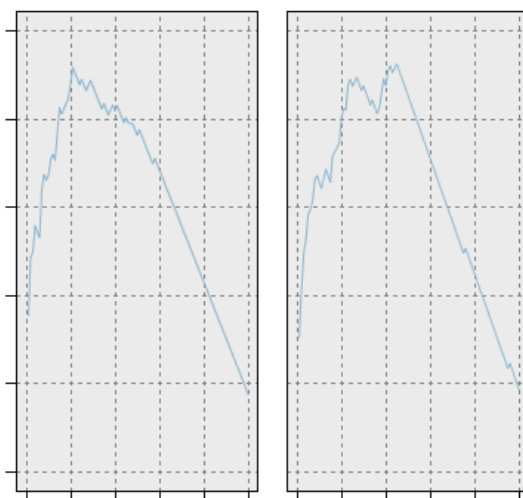
Testing the model

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.For example, you can look at a handful of customer records and see which customers are likely to respond to a direct mail campaign.

▶ On the Modeling tab, click Test; the Test dialog is displayed.

▶ Select the data source you want to use, or select Specify custom data in the Records section to enter test records manually.

▶ 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. For more information, see the topic Defining selection rules in Chapter 5 on p. 39.

▶ 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. If testing an association model using transactional data, input data is sampled at the transactions level, not the record level.

▶ 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.

▶ Under Select Fields to Display, select the fields you want to appear in the test output.

▶ Click Run.

▶ The test summary displays a list of successful allocations for each records.

▶ Click the information icon at the beginning of each row to obtain details for that row.

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. For more information, see the topic Creating rules in Chapter 5 on p. 37.

■ For applications that support batch scoring, model scores can be written to a database table or flat file.For more information, see the topic Scoring models to a database table, file, or Cognos BI server in Chapter 7 on p. 62.

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. For more information, see the topic Measurement levels in Chapter 4 on p. 27.

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. For more information, see the topic Building a predictive rule model on p. 45.

- **$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.

Scoring Models

Scoring models to a database table, file, or Cognos BI server

To score a model means to apply it to the data sample or population you are interested in. For example, you might use your existing customers to generate a model that predicts which customers are most likely to respond to a campaign, and then use the model to score records for people who are not currently your customers, but whom you might want to include in a promotional mailing.

The basic steps in scoring a model are:

▶ Select the destination for the scored data.

▶ Select the information to be output and how it is to be stored.

▶ Select the data source, and any subsets of that data, to be scored.

▶ Specify which records are to be scored, and preview the expected results.

▶ Click Score Now to run scoring against the model using your settings.

Scores are written to one or more fields which are added to the data. By convention, the names of these fields are based on the target field, but with a prefix added. For example, if the target field is named *Response*, the output fields will have names like *$XF-Response* and *$XFRP-Response*. For more information, see the topic Selecting output fields on p. 65.

Selecting the scoring destination

You can choose to output the scoring data to a database, file, or Cognos BI server. Supported file types include text (**.txt*, **.csv*), Excel, the **.sav* format used by IBM® SPSS® Statistics, SAS, and those supported by IBM® SPSS® Data Collection. For more information, see the topic File destination on p. 63.

Database destination

When you choose to send scoring output to a database you are prompted to enter further details.

▶ Enter the Database Name or browse for the one you require.

▶ Specify how scores are to be stored. You can choose to create a new table, in which case you are prompted to enter a name for the table. Alternatively, you can either append them to an existing table or overwrite an existing data source, in which case you are prompted to select the relevant table or data source.

When you append to, or overwrite, an existing database you can map the scoring fields to fields in the existing destination. For more information, see the topic Map scoring fields on p. 65.

File destination

Scores can be output to several different file types. For each type, you either browse for an existing file or create a new file to which the scoring data is added.

When you append to, or overwrite, an existing file you can map the scoring fields to fields in the existing destination. For more information, see the topic Map scoring fields on p. 65.

Flat file

When you choose to send scoring output to a flat file you can enter further details to control the file contents.

Write mode. If Overwrite is selected, any existing data in the specified file will be overwritten. If Append is selected, output will be added to the end of the existing file, preserving any data it contains.

■ **Include field names.** If this option is selected, field names will be written to the first line of the output file. This option is available only for the Overwrite write mode.

New line after each record. If this option is selected, each record will be written on a new line in the output file.

Field separator. Specifies the character to insert between field values in the generated text file. Options are Comma, Tab, Space, and Other. If you select Other, enter the desired delimiter character(s) in the text box.

Symbol quotes. Specifies the type of quoting to use for values of symbolic fields. Options are None (values are not quoted), Single ('), Double ("), and Other. If you select Other, enter the desired quoting character(s) in the text box.

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.

Excel files

Type. Select the Excel file type that you are exporting.

Include field names. Specifies whether field names should be included in the first row of the worksheet.

SAS files

Type. Select the SAS file format to be created. You can choose from three SAS file formats: SAS for Windows/OS2 (*.sd2), SAS for UNIX (*.ssd), or SAS Version 7/8/9 (*sas7bdat).

Export field names. Select options for exporting field names and labels for use with SAS.

- ■ **Names and variable labels.** Select to export both field names and field labels. Names are exported as SAS variable names, while labels are exported as SAS variable labels.

- ■ **Names as variable labels.** Select to use the field names as variable labels in SAS. To prevent possibly creating invalid SAS names, select Names and variable labels instead.

IBM SPSS Statistics files

Export field names. Specifies a method of handling variable names and labels upon export to a IBM® SPSS® Statistics *.sav* file. Note that any invalid names are automatically corrected by replacing invalid characters.

- ■ **Names and variable labels.** Select to export both field names and field labels. Names are exported as SPSS Statistics variable names, while labels are exported as SPSS Statistics variable labels.

- ■ **Names as labels.** Select to use the field names as variable labels in SPSS Statistics. To prevent possibly creating invalid SPSS Statistics names, select Names and variable labels instead.

Cognos BI server destination

When you choose to send scoring output to a Cognos BI server you are prompted to enter further details.

Note: You can export only relational data, not OLAP data.

To export data to Cognos BI, you must specify the following:

- ■ **Cognos connection.** The connection to the Cognos BI server.
- ■ **ODBC connection.** The connection to the Cognos data server that the Cognos BI server uses.

The connections must point to the same database, and the username and password for the Cognos server to connect to ODBC must be the same as the ODBC details.

You export the actual data to the data server, and the package metadata to the Cognos BI server.

▶ Select the Cognos BI Server Destination and click Connect; you are prompted to enter the server connection details.

▶ 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.

▶ 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.

▶ 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.

▶ Select the Data source.

▶ Type the path and name of the Folder on the Cognos BI server where the export package is to be created. Alternatively, you can browse for the one you require.

▶ Type the Package name of the package that is to contain the exported metadata. This must be a new package; you cannot export to an existing package. For more information, see the topic Selecting Cognos object details in Chapter 4 on p. 32.

▶ To select the ODBC Destination, enter the Database name or browse for the one you require.

▶ Specify how scores are to be stored. You can choose to create a new table, in which case you are prompted to enter a name for the table. Alternatively, you can either append them to an existing table or overwrite an existing data source, in which case you are prompted to select the relevant table or data source.

When you append to, or overwrite, an existing database you can map the scoring fields to fields in the existing destination. For more information, see the topic Map scoring fields on p. 65.

Map scoring fields

When you append or overwrite scores to an existing database table or file, you may need to map the scoring fields to fields in the existing table or file. Note that the output and associated mapped field must have the same data type.

■ When writing scores to an existing database table, all fields must be mapped to existing fields in order for scoring to proceed, whether overwriting or appending the new scores. If necessary, click Map Fields to specify outputs for any fields that remain.

■ When writing to an existing file, mapping is only necessary if Append is selected under output file options. If overwriting an existing file, mapping is not required as the new file will simply replace the old one.

▶ 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.

▶ Select from the available output destinations and click Map to map them to the required scoring fields.

▶ When finished, click Save.

Selecting output fields

On the Score tab, select Specify output fields to choose the fields to include with the scoring output. In addition to the actual scores, you will typically want at least one field that allows you to identify each record, such as a customer id field. Optionally, some or all of the input fields used in modeling may also be included.

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. For more information, see the topic Measurement levels in Chapter 4 on p. 27.

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. For more information, see the topic Building a predictive rule model in Chapter 6 on p. 45.
- **$DI-<target>.** Index number used to identify the segment into which the record falls (displayed in the left-most column in the modeling results).
- **$DRP-<target>.** Propensity score indicating the likelihood of the desired response. Available for flag targets only.

Selecting data and subsets

From the Data source to score list, select the required data source.

Data subsets

If you don't want to score the entire data source, you can select a subset of the data by using an existing rule, or creating a new rule that will only be used for this model. For more information, see the topic Defining selection rules in Chapter 5 on p. 39.

Selecting records to be scored

For models that return propensity scores, you can use these values to select the records to be scored. For example, you could score the 30% of customers with the highest propensity to respond, and use these to produce a mailing list. Propensity scores indicate the likelihood of the "true" value for the target, and are available for predictive models with flag targets, as well as predictive rule (interactive) models. (In the case of predictive rule models, the selected response is the "true" value and all others are false. For other types of models, these options are not available.)

There are three ways you can choose to select records based on propensity:

- **The top percentage.** By default this selects the top ten percent. This option enables you to simulate your scoring results based on profit. For more information, see the topic Simulating maximum profit in Chapter 6 on p. 57.
- **The top number of records.** By default this selects the top 1000 records, where available.
- **By specifying minimum and maximum propensity values.** For example, all records that have a propensity to respond between 75% and 100%.

Scoring models

Once you have specified your scoring details, you can preview a sample of the model scores. This is especially useful when you have a large dataset since it enables you to quickly check that the scoring results appear as you expect, without having to wait for the entire dataset to be scored.

Note: If the data is found to be invalid or incomplete, a warning is displayed when you attempt to preview it.

The preview shows data as a table that displays records by row and variables by columns.

Scoring

To establish how many records will be scored, click the refresh icon at the bottom of the screen. This provides a count of the number of records that will be used to build the model, taking into account the total number of records in the data source and any optional settings you specify.

When you click Score Now, a progress bar is displayed across the bottom of the tab. You can stop scoring at any time while this is displayed.

Note: If the data is found to be invalid or incomplete, a warning is displayed when you attempt to score it.

Once the data is successfully scored a message informs you where the results have been stored - either in the database or file you specified.

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.

► Open Deployment Manager.

► From the Tools menu, select Server Administration.

► On the Server Administration tab, connect to your server.

► Under Users and Groups, select Local User Repository.

► 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.

► 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 8-1
Role actions

Role	Actions Required
Decision Management administrator	Access Contents and Folders Decision Management Administration Run Report Dynamically Score Model Submit Work Show All Versions
Decision Management user	Access Contents and Folders 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.

▶ Select the Decision Management administrator role and click Edit Users and Groups. Add the Decision Management Administrators group you just created.

▶ 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:

▶ Select In default launcher for applications you want all users to see on their launch page.

▶ 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. For more information, see the topic Launching applications in Chapter 3 on p. 17.

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.

▶ On the Home page, select Hide Application home page.

▶ 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. For more information, see the topic Managing data sources in Chapter 4 on p. 24.

Locking scoring destination options

For applications configured to use the Score Now section (Deploy or Score tab), there is a section that allows business users to specify the destination for storing score data (to database or file). For example, if storing scores in a database, you specify the database name and the method for storing the scores (such as creating a new table in the database).

As an administrator, you can lock the destination options for scoring. For example, you may want scores to only be appended to an existing table in a certain database—instead of giving business users the option to overwrite existing data, append new data to an existing table, create a new table, or store data in a file.

► To lock all scoring destination options, click the padlock icon next to the Destination drop-down on the Deploy tab. Business users will be able to see the destination options, but not edit them.

Downloading project metadata for use in creating external rules

Rules created in a Business Rules Management System such as ILOG 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, the IBM Analytical Decision Management administrator can download a *.ZIP* file with metadata for the current project. The downloaded file is saved locally on the system used by the administrator to trigger the download.

Figure 8-1
Download metadata icon

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 ILOG Rules Studio 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.

Note that it is possible to create and utilize ILOG 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* dataresponse rule, the XML schema definition (*.XSD*) 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 dataresponse 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 ILOG rules projects, and one representing a rules return type which should be used for the output parameter of ILOG rules projects. 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 ILOG 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>


I   <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>
```

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

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

Documentation is included on the installation disc for each product or application, or
provided in electronic download files. Documentation is installed into the IBM® SPSS®
Collaboration and Deployment Services installation directory (for example, *C:\Program
Files\IBM\SPSS\Deployment\5.0\help\en\DecisionManagement*).

PDF documentation is also available on the Web at
http://www-01.ibm.com/support/docview.wss?uid=swg27024118, and release notes are available
at http://www-01.ibm.com/support/docview.wss?uid=swg27024117

For an online, centralized collection of all documentation, see the IBM Analytical Decision
Management 7 Information Center. The following table provides links to each deliverable in the
Information Center.

Table 9-1
IBM Analytical Decision Management documentation

Documentation	Description
Installation Guide for Windows Installation Guide for UNIX	Instructions for installing and launching the product.
Application Designer's Guide	Instructions for designing and configuring applications, including a complete XML schema reference, as well as details on customizing the user interface.
Application User's Guide	Includes general information about the user interface, intended for business users.

Documentation	Description
IBM SPSS Modeler Advantage	IBM® SPSS® Modeler Advantage puts the power of predictive modeling in the hands of business users. Using predictive models, you can identify patterns based on what has happened in the past, and use them to predict what is likely to happen in the future.
IBM Analytical Decision Management for Customer Interactions	IBM® Analytical Decision Management for Customer Interactions decides which promotions to offer customers when they contact your organization, and delivers recommendations to a call center, web site, or store location in real time. By combining the logic of business rules with the insight gained through predictive modeling, the application identifies the most profitable decision for each customer.
IBM Analytical Decision Management for Campaign Optimization	IBM® Analytical Decision Management for Campaign Optimization is similar to IBM Analytical Decision Management for Customer Interactions . But IBM Analytical Decision Management for Campaign Optimization uses optimization and has the added value of maximizing the return on your campaign investment by determining the best offers for individual customers within constraints such as your budget, channel capacity, and contact policies.
IBM Analytical Decision Management for Claims	IBM® Analytical Decision Management for Claims combines the logic of business rules with the insight gained through predictive models, allowing organizations to process incoming claims in real time and determine the best action for each. For example, claims can be set on a "fast track" for immediate payment, processed in the normal manner, or referred to the special investigations unit.
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.

Sharing streams

Models and projects created in IBM® Analytical Decision Management are saved as stream files, and can be modified and used in IBM® SPSS® Modeler if available. For example, you may want to create a model or project stream using IBM® SPSS® Modeler Advantage or another IBM Analytical Decision Management application, then customize the stream in SPSS Modeler before deploying through IBM Analytical Decision Management. Alternatively, streams created in SPSS Modeler can be opened in IBM Analytical Decision Management, provided deployment settings have been specified as described below.

Opening streams in IBM Analytical Decision Management

When opening streams in IBM Analytical Decision Management or IBM SPSS Modeler Advantage:

- If deployment type is not specified (none), the stream can be opened but only the Data tab is available.

- If deployment type is Scoring Only, the Score tab may also be available, provided the scoring branch starts with a single data source node of a type supported by IBM SPSS Modeler Advantage or IBM Analytical Decision Management (no merged data sets).

- If deployment type is Model Refresh, the Modeling tab may also be available, provided the modeling branch starts with a single data source node of a type supported by IBM SPSS Modeler Advantage or IBM Analytical Decision Management, and the selected modeling node is either an automated modeling node (supported for predictive modeling) or a Decision List node (supported for predictive rule model building). In addition, a Type node must be present that defines the target field as well as other information used in modeling.

Specifying deployment settings in IBM SPSS Modeler

In SPSS Modeler, deployment settings are specified in the Stream Properties dialog box, as follows:

▶ From the SPSS Modeler menus, choose:
Tools > Stream Properties > Deployment

▶ From the Deployment type list, choose Scoring Only or Model Refresh as appropriate.

▶ Select a Scoring node, Modeling node, and Model node as needed.

Select Check to confirm that all required options have been specified, or select Store to store the stream in IBM® SPSS® Collaboration and Deployment Services Repository.

For streams created in IBM SPSS Modeler Advantage or IBM Analytical Decision Management, deployment information is set automatically.

Tips for using IBM SPSS Modeler streams in IBM Analytical Decision Management

■ It may be easier to create streams using IBM SPSS Modeler Advantage or IBM Analytical Decision Management and then modify in SPSS Modeler, rather than building from scratch using SPSS Modeler. Streams created in this manner will be structured in a manner that allows them to be fully functional in IBM Analytical Decision Management.

■ For all streams created in IBM SPSS Modeler Advantage, supernodes are included at certain locations within the branch that can be customized by an expert user. It is strongly recommended that any customizations or additions to the stream be added within these supernodes to ensure they are preserved in the stream and recognized by IBM SPSS Modeler Advantage.

■ Any local rules nodes containing an expression rule immediately downstream of each source node will have their fields included in the Expression Manager. Note that local rules can only be created within IBM Analytical Decision Management or IBM SPSS Modeler Advantage and not by SPSS Modeler.

■ If there is a Partition node upstream of the modeling node (and downstream of the Type node) then Automatically partition data to enable model evaluation on build data source for evaluation and testing is enabled in IBM SPSS Modeler Advantage, and the check box state is set depending on whether the partition node is enabled or not.

■ If there is an Automated Data Preparation (ADP) node downstream of the Type node, then Automatically clean up and prepare data for reliable model building is enabled and the check box state is set depending on whether the ADP node is enabled or not.

■ If Rule or Select nodes are present upstream of the modeling node, these are used to populate the Specify selections to use section.

■ If stream parameters have been defined (Parameters tab in the Stream Properties dialog box), a Parameters link will be displayed on the Modeling and Score tabs in IBM SPSS Modeler Advantage. The user can click this link to specify these values.

Launching IBM SPSS Modeler Advantage from IBM SPSS Modeler

From the SPSS Modeler splash screen, you can choose to create a model with IBM SPSS Modeler Advantage.

Accessibility

Accessibility features help sers 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.

Help accessibility

An alternate version of the user help system installed with the product is available to provide additional support for screen readers. To switch to the alternate version you will need to run a batch file to convert the existing help system into a format more efficiently read by a screen reader. In the alternate version the table of contents is on the right and help topics display on the left.

To enable support for screen readers

▶ Navigate to the root directory of the help on the server (for example, *C:\Program Files\IBM\SPSS\Deployment\5.0\help\en\DecisionManagement\ClaimsManagment\userhelp*).

▶ To switch to the screen reader-friendly format double-click the batch file *format-for-screen-readers.bat* to run it.

To switch back to the default format run the batch file *restore-default-format.bat*.

Note: The IBM Analytical Decision Management 7 Information Center is accessibility-enabled.

Notices

This information was developed for products and services offered worldwide.

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 character set (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 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 wish 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, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

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, ibm.com, and SPSS are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at *http://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.

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 of Sun Microsystems, Inc. in the United States, other countries, or both.

SAS is a registered trademark of SAS Institute Inc. in the United States, other countries, or both.

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

Index