Oracle Classification Assignment – Tutorial

Note: This is a slightly edited version of Oracle's tutorial on how to use Oracle Data Miner. The original tutorial can be found from the following line:

http://www.oracle.com/webfolder/technetwork/tutorials/obe/db/11g/r2/prod/bidw/datamining/ODM11gR2.htm

Note: There are seven "**Required"** screens.

Purpose

This tutorial covers the use of Oracle Data Miner to perform data mining against Oracle Database 11*g* Release 2. In this lesson, you examine and solve a data mining business problem by using the Oracle Data Miner graphical user interface (GUI). The Oracle Data Miner GUI is included as an extension of Oracle SQL Developer, version 3.0.

Oracle SQL Developer is a free graphical tool for database development. With SQL Developer, you can browse database objects, run SQL statements and SQL scripts, and edit and debug PL/SQL statements. Starting with SQL Developer, version 3.0, you can also access the Oracle Data Miner GUI, which provides a tightly integrated interface to Oracle Data Mining features.

Please not that some screens in this tutorial are not exactly the same as the ones you will see from your virtual machine.

The whole tutorial will take an hour or two. You may want to proceed slowly to understand what each step does.

Overview

Data mining is the process of extracting useful information from masses of data by extracting patterns and trends from the data. Data mining can be used to solve many kinds of business problems, including:

- Predict individual behavior, for example, the customers likely to respond to a promotional offer or the customers likely to buy a specific product (Classification)
- Find profiles of targeted people or items (Classification using Decision Trees)
- Find natural segments or clusters (Clustering)
- Identify factors more associated with a target attribute (Attribute Importance)
- Find co-occurring events or purchases (Associations, sometimes known as Market Basket Analysis)
- Find fraudulent or rare events (Anomaly Detection)

The phases of solving a business problem using Oracle Data Mining are as follows:

1. Problem Definition in Terms of Data Mining and Business Goals
2. Data Acquisition and Preparation
3. Building and Evaluation of Models
4. Deployment

Problem Definition and Business Goals

When performing data mining, the business problem must be well-defined and stated in terms of data mining functionality. For example, retail businesses, telephone companies, financial institutions, and other types of enterprises are interested in customer “churn” – that is, the act of a previously loyal customer in switching to a rival vendor.

The statement “I want to use data mining to solve my churn problem” is much too vague. From a business point of view, the reality is that it is much more difficult and costly to try to win a defected customer back than to prevent a disaffected customer from leaving; furthermore, you may not be interested in retaining a low-value customer. Thus, from a data mining point of view, the problem is to predict which customers are likely to churn with high probability, and also to predict which of those are potentially high-value customers.

Data Acquisition and Preparation

A general rule of thumb in data mining is to gather as much information as possible about each individual, then let the data mining operations indicate any filtering of the data that might be beneficial. In particular, you should not eliminate some attribute because you think that it might not be important – let ODM’s algorithms make that decision. Moreover, since the goal is to build a profile of behavior that can be applied to any individual, you should eliminate specific identifiers such as name, street address, telephone number, etc. (however, attributes that indicate a general location without identifying a specific individual, such as Postal Code, may be helpful.)

It is generally agreed that the data gathering and preparation phase consumes more than 50% of the time and effort of a data mining project.

Building and Evaluation of Models

The Workflow creation process of Oracle Data Miner automates many of the difficult tasks during the building and testing of models. It’s difficult to know in advance which algorithms will best solve the business problem, so normally several models are created and tested.

No model is perfect, and the search for the best predictive model is not necessarily a question of determining the model with the highest accuracy, but rather a question of determining the types of errors that are tolerable in view of the business goals.

Deployment

Oracle Data Mining produces actionable results, but the results are not useful unless they can be placed into the correct hands quickly. The Oracle Data Miner user interface provides several options for publishing the results.

Scenario

This lesson focuses on a business problem that can be solved by applying Classification models. In our scenario, ABC Company wants to identify customers who are most likely to purchase insurance.

Note: For the purposes of this tutorial, the "Data and Acquisition" phase has already been completed, and the sample data set contains all required data fields. Therefore, this lesson focuses primarily on the "Building and Evaluation of Models" phase.

Create a Data Miner Project

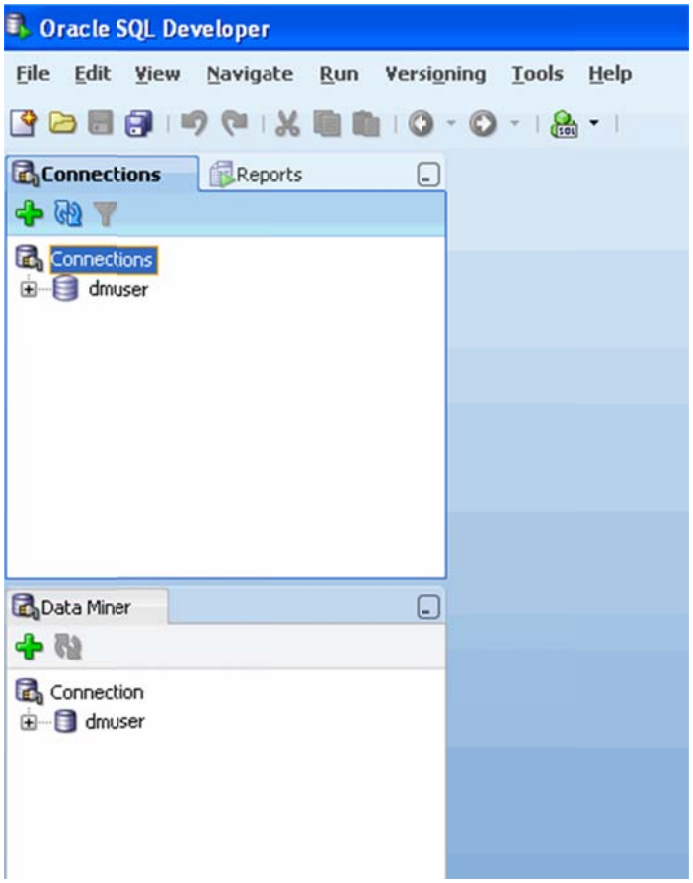
Before you create a Data Miner Project and build a Data Miner workflow, it is helpful to examine some of the Data Miner interface components within SQL Developer. You can then structure your working environment to provide simplified access to the necesary Data Miner features.

Identifying SQL Developer and Data Miner Interface Components

After setting up Oracle Data Miner for use within SQL Developer, different interface elements may be displayed, including both SQL Developer and Data Miner components.

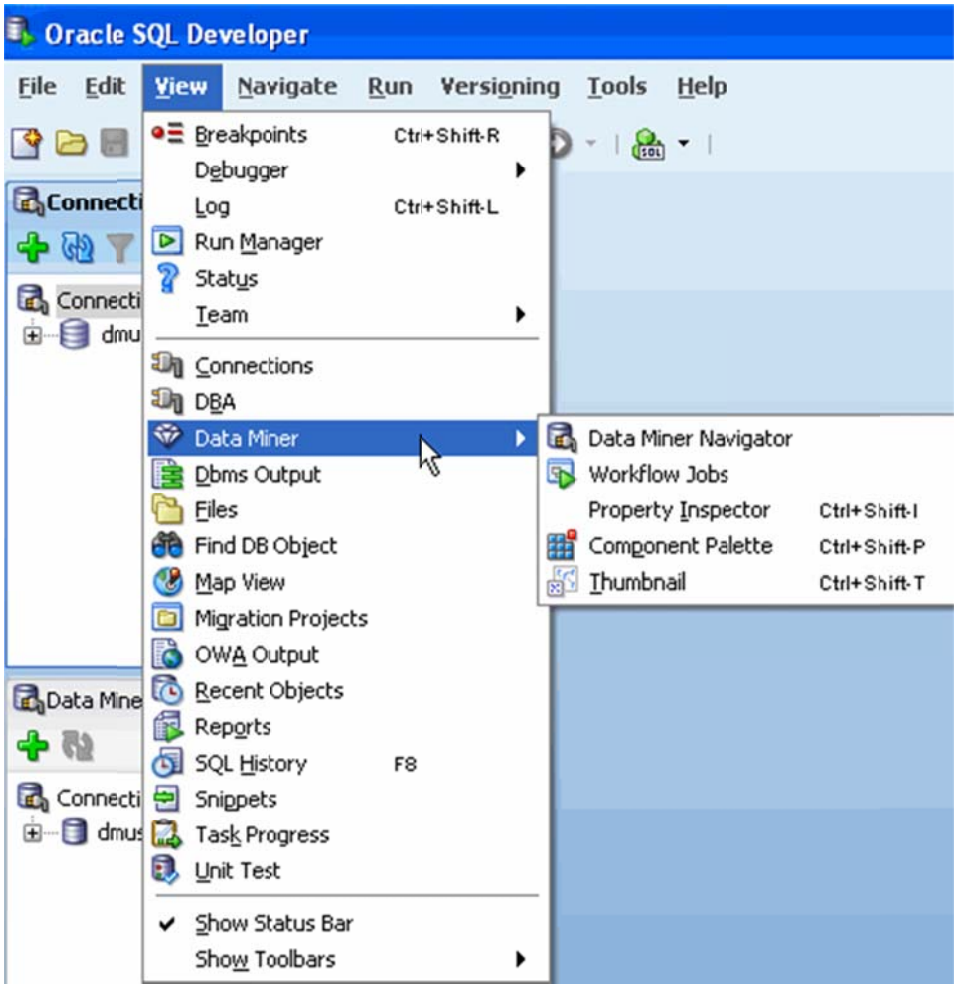
In the following example, several display elements are open, including the:

- The SQL Developer Connections tabs
- The SQL Developer Reports tabs
- The Data Miner tab



Notes:

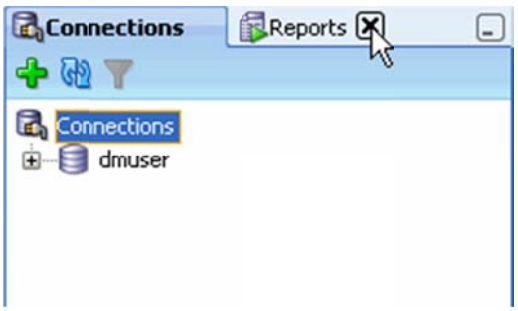
- The layout and contents of your SQL Developer window may be different than the example shown above.
- SQL Developer and Data Miner interface elements open automatically when needed.
- Additional Data Miner interface elements include the Workflow Jobs, Property Inspector, Component Palette, and Thumbnail tabs. You can open any one them manually from the main menu by using **View** > **Data Miner** > as shown here:

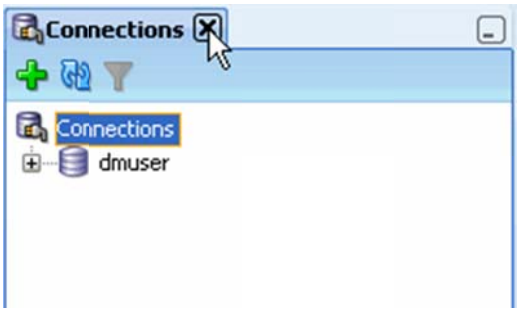


In order to simplify the interface for Data Mining development, you can dismiss the SQL Developer specific interface elements by clicking on the respective Close **[x]**icons for each tab or window.

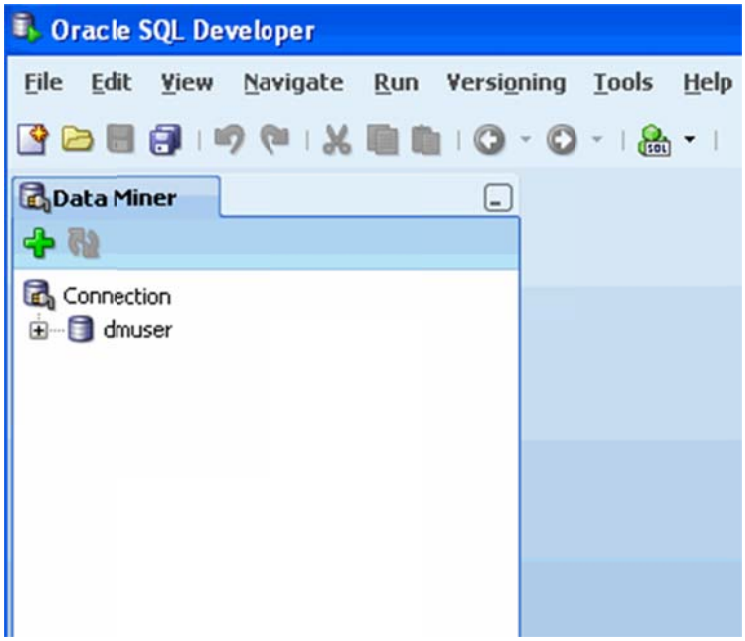
For example, close both of the SQL Developer tabs mentioned above:

1. SQL Developer **Reports** tab
2. SQL Developer **Connections** tab

1. 

2. 

Now, the SQL Developer interface should look like this:



Before creating your first project, ensure that the Workflow Jobs tab is also open. To do this, select **View > Data Miner > Workflow Jobs**.

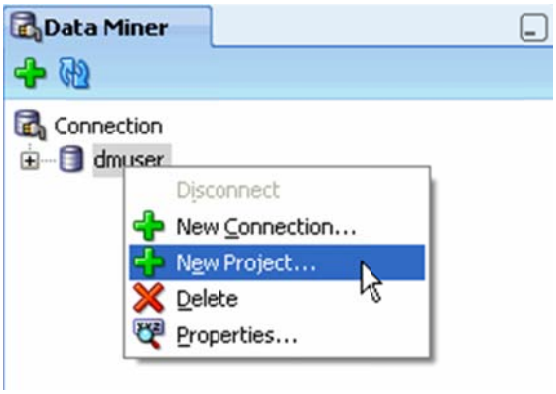
Note: You can re-open the SQL Developer Connections tab (and other interface elements) at any time by using the View menu.

Create a Data Miner Project

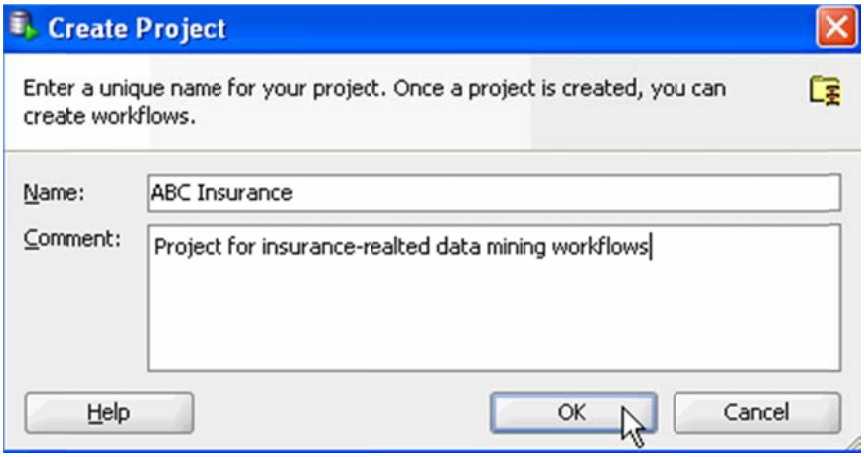
Before you begin working on a Data Miner Workflow, you must create a Data Miner Project, which serves as a container for one or more Workflows.

To create a Data Miner Project, perform the following steps:

1. In the Data Miner tab, right-click the dmuser connection that you previously created, and select **New Project**, as shown here:

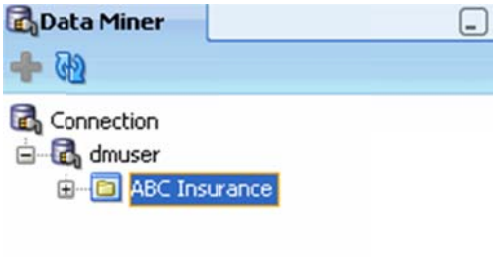


2. In the Create Project window, enter a project name (in this example ABC Insurance) and then click **OK**.



Note: You may optionally enter a comment that describes the intentions for this project. This description can be modified at any time.

Result: The new project appears below the data mining user connection node.



Build a Data Miner Workflow

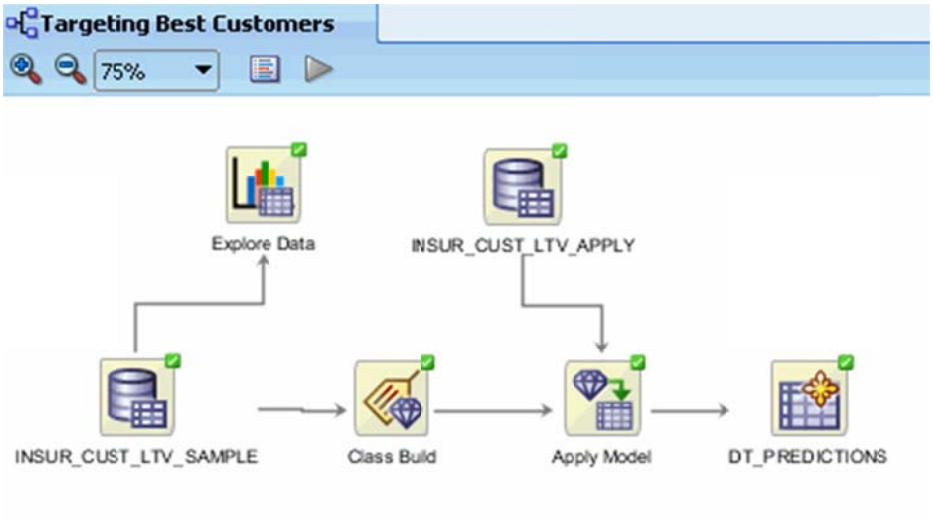
A Data Miner Workflow is a collection of connected nodes that describe a data mining processes.

A workflow:

- Provides directions for the Data Mining server. For example, the workflow says "Build a model with these characteristics." The model is built by the data mining server with the results returned to the workflow.
- Enables you to interactively build, analyze, and test a data mining process within a graphical environment.
- May be used to test and analyze only one cycle within a particular phase of a larger process, or it may encapsulate all phases of a process designed to solve a particular business problem.

What Does a Data Miner Workflow Contain?

Visually, the workflow window serves as a canvas on which you build the graphical representation of a data mining process flow, like the one shown here:



Notes:

- Each element in the process is represented by a graphical icon called a *node*.
- Each node has a specific purpose, contains specific instructions, and may be modified individually in numerous ways.
- When linked together, workflow nodes construct the modeling process by which your particular data mining problem is solved.

As you will learn, any node may be added to a workflow by simply dragging and dropping it onto the workflow area. Each node contains a set of default properties. You modify the properties as desired until you are ready to move onto the next step in the process.

Sample Data Mining Scenario

In this topic, you will create a data mining process that predicts which existing customers are most likely to purchase insurance.

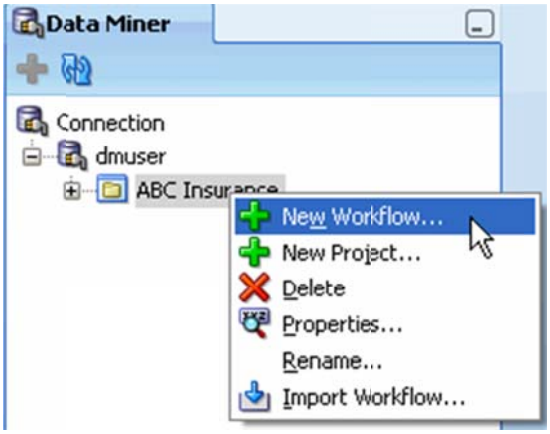
To accomplish this goal, you build a workflow that enables you to:

- Identify and examine the source data
- Build and compare several Classification models
- Select and run the models that produce the most actionable results

To create the workflow for this process, perform the following steps.

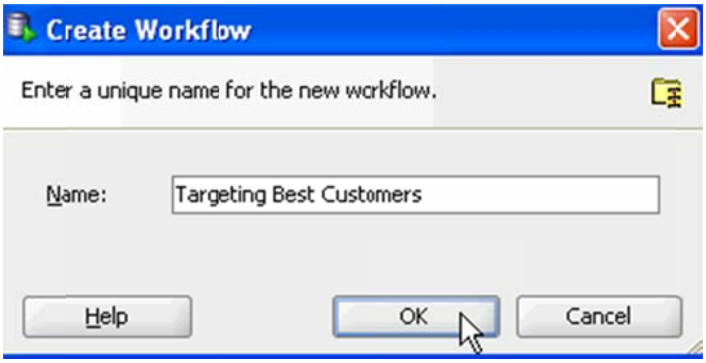
Create a Workflow and Add a Data Source

1. Right-click your project (ABC Insurance) and select **New Workflow** from the menu.



Result: The Create Workflow window appears.

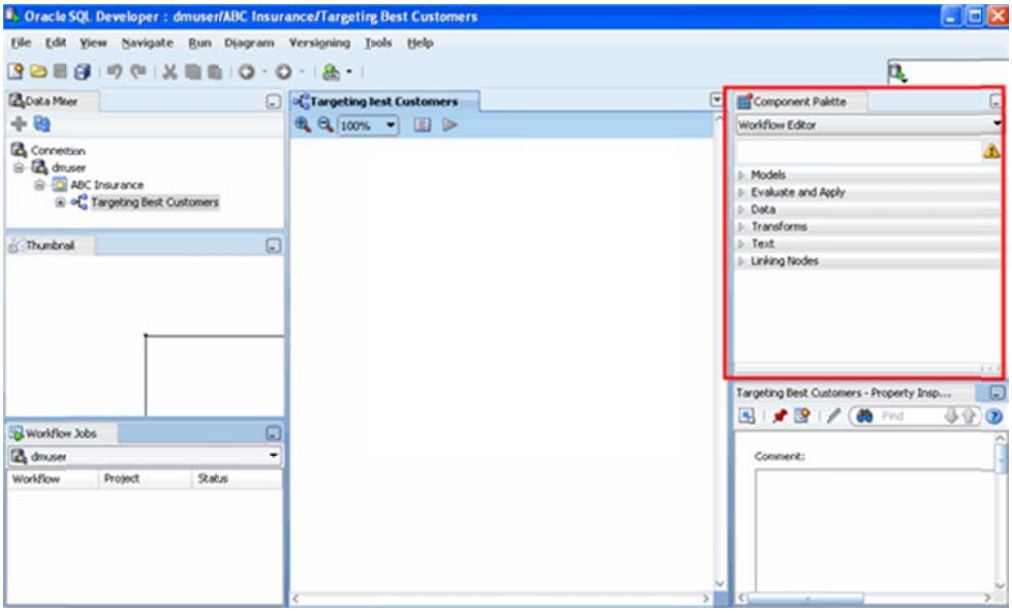
2. In the Create Workflow window, enter **Targeting Best Customers** as the name and click **OK**.



Result:

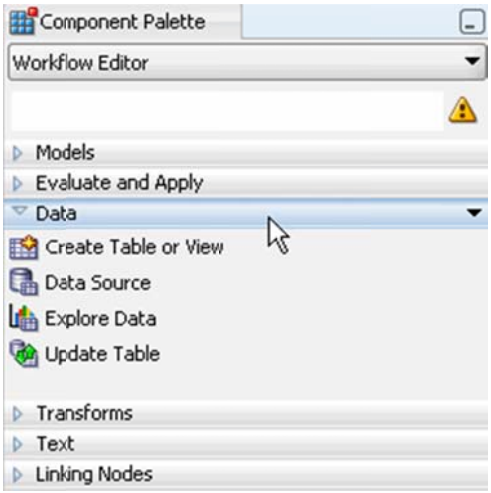
In the middle of the SQL Developer window, an empty workflow canvas opens with the name that you specified.
On the right-hand side of the interface, the Component Palette tab of the Workflow Editor appears (shown below with a red border).
In addition, three other Oracle Data Miner interface elements are opened:

- The Thumbnail tab
- The Workflow Jobs tab
- The Property Inspector tab



3. The first element of any workflow is the source data. Here, you add a Data Source node to the workflow, and select the INSUR_CUST_LTV_SAMPLE table as the data source.

A. In the Component Palette, click the **Data** category. A list of data nodes appear, as shown here:

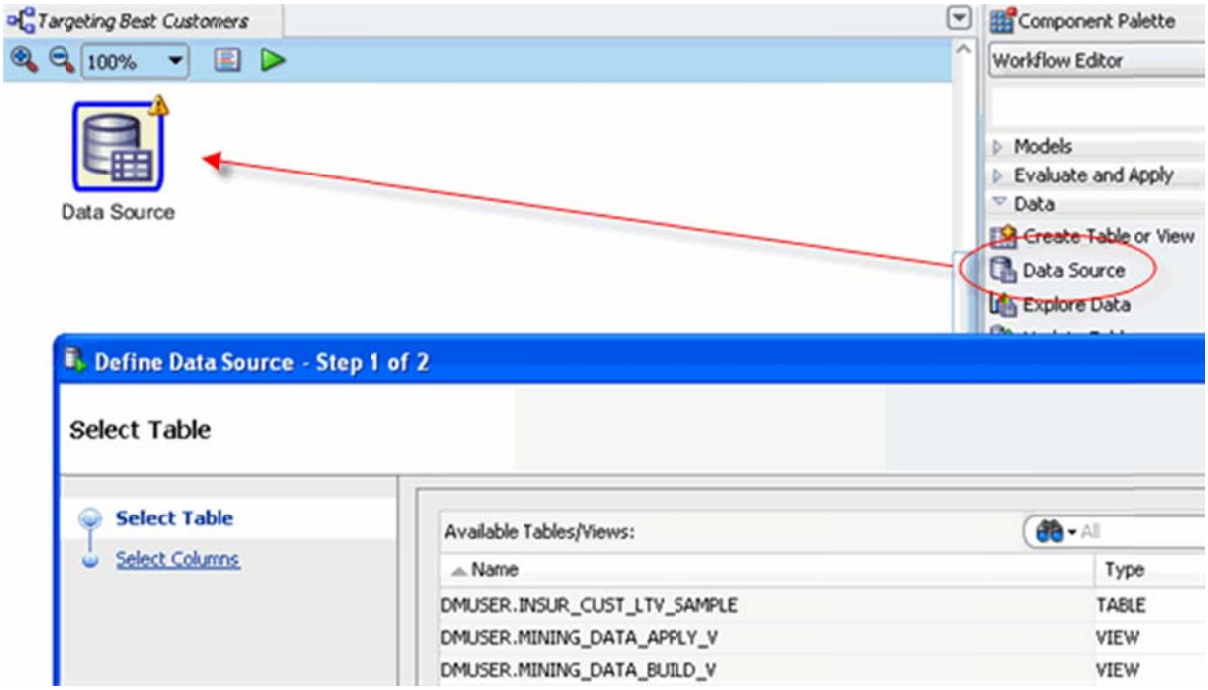


B. Drag and drop the **Data Source** node onto the Workflow pane.

Result: A Data Source node appears in the Workflow pane and the Define Data Source wizard opens.

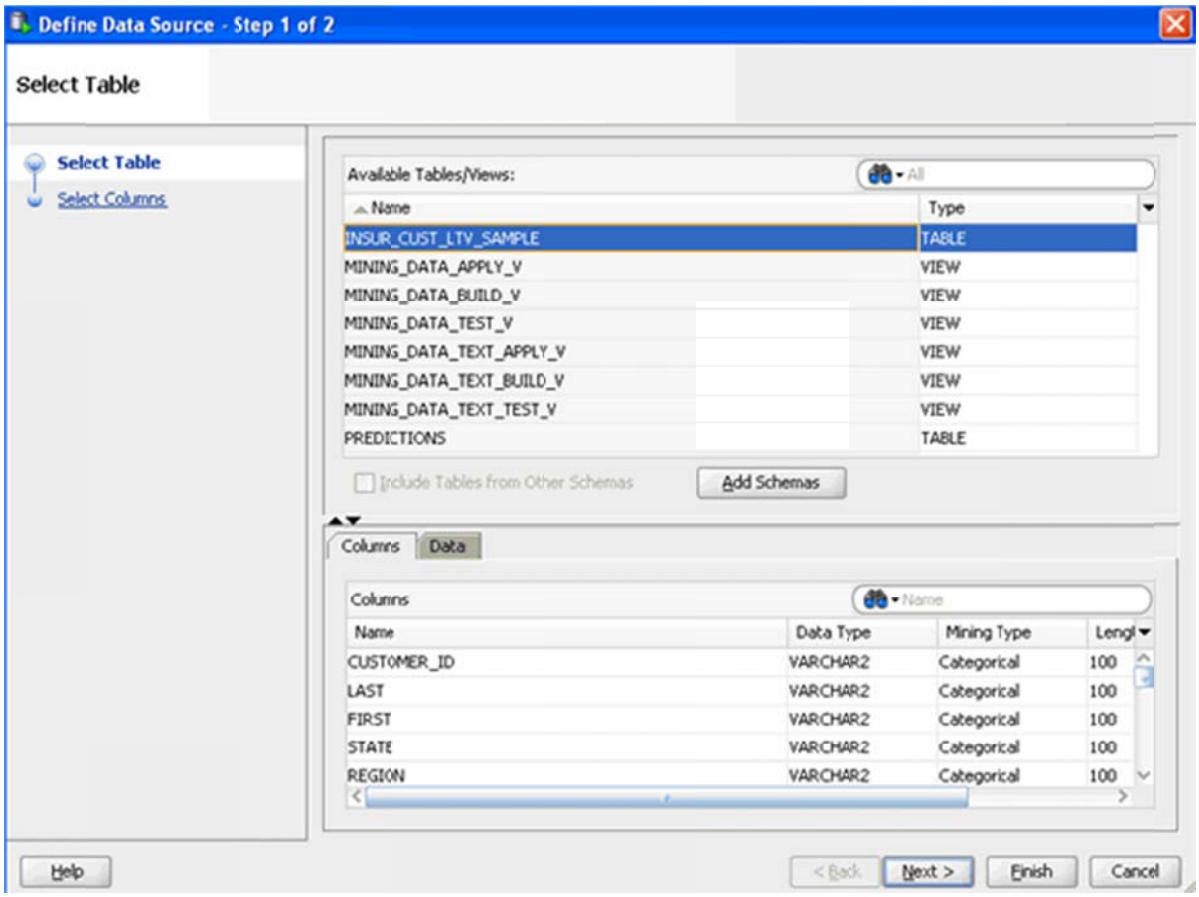
Notes:

- Workspace node names and model names are generated automatically by Oracle Data Miner. In this example, the name "Data Source" is generated. You may not get exactly the same node and model names as shown in this lesson.
- You can change the name of any workspace node or model using the Property Inspector.



4 . In Step 1 of the wizard:

A. Select **INSUR_CUST_LTV_SAMPLE** from the Available Tables/Views list, as shown here:

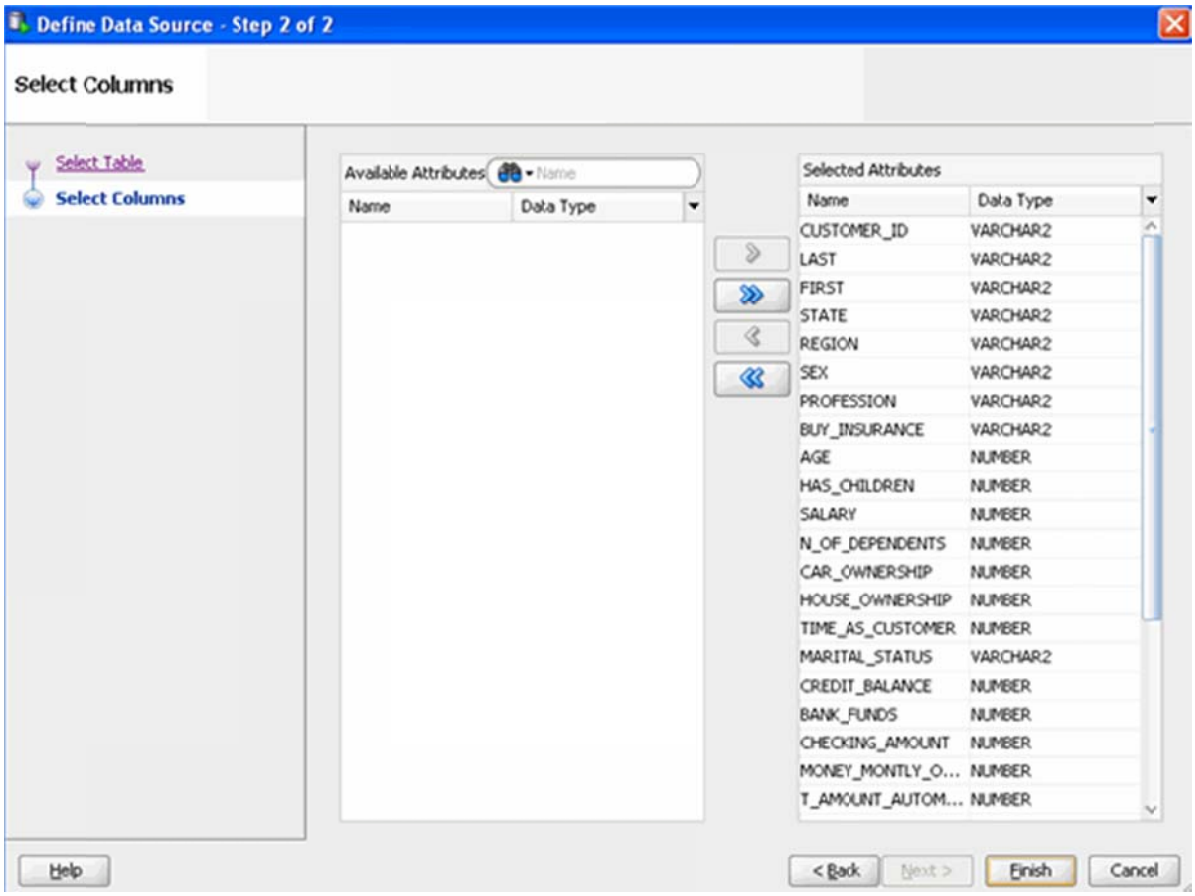


Note: You may use the two tabs in the bottom pane in the wizard to view and examine the selected table. The Columns tab displays information about the table structure, and the Data tab shows a subset of data from the selected table or view.

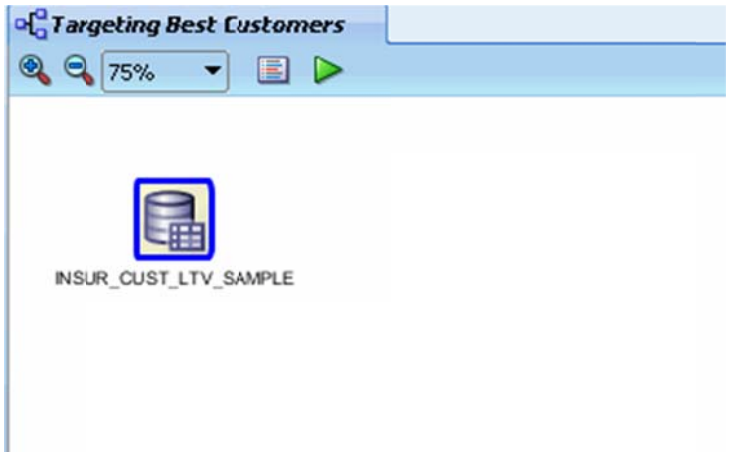
B. Click **Next** to continue.

5 . In Step 2 of the wizard, you may remove individual columns that you don't need in your data source. In our case, we'll keep all of the attributes that are defined in the table.

At the bottom of the wizard window, click **Finish**.



Result: As shown below, the data source node name is updated with the selected table name, and the properties associated with the node are displayed in the Property Inspector, located below the Component Palette pane.



Notes:

You can resize nodes in the workflow canvas by entering or selecting a different value from the Zoom options. Notice that **75%** has been selected from the Zoom pull-down list.

You can add descriptive information about any node by using the Details tab in the Property Inspector.

The Thumbnail tab also provides a smaller display of the larger workflow window. As you drag nodes around the workflow window, the thumbnail view automatically adjusts.

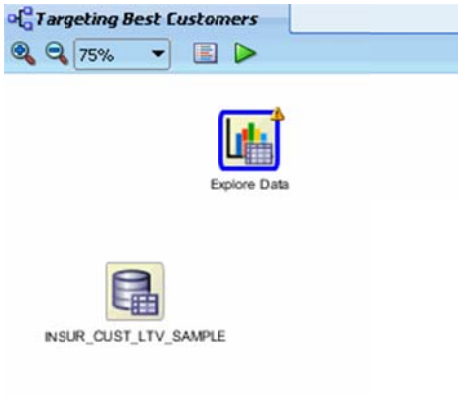
Examine the Source Data

You can use an Explore Data node to examine the source data. Although this is an optional step, Oracle Data Miner provides this tool to enable you to verify if the selected data meets the criteria to solve the stated business problem.

Follow these steps:

1. Drag and drop the Explore Data node from the Component Palette to the Workflow, like this:

Result: A new Explore Data node appears in the workflow pane, as shown here. (As before, a node name is automatically generated.)



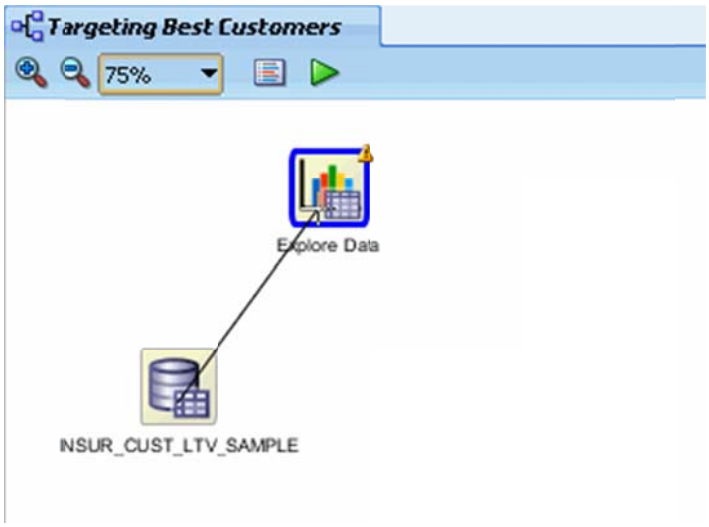
Notes:

A yellow Information (!) icon in the border around any node indicates that it is not complete. Therefore, at least one addition step is required before the Explore Data node can be used.

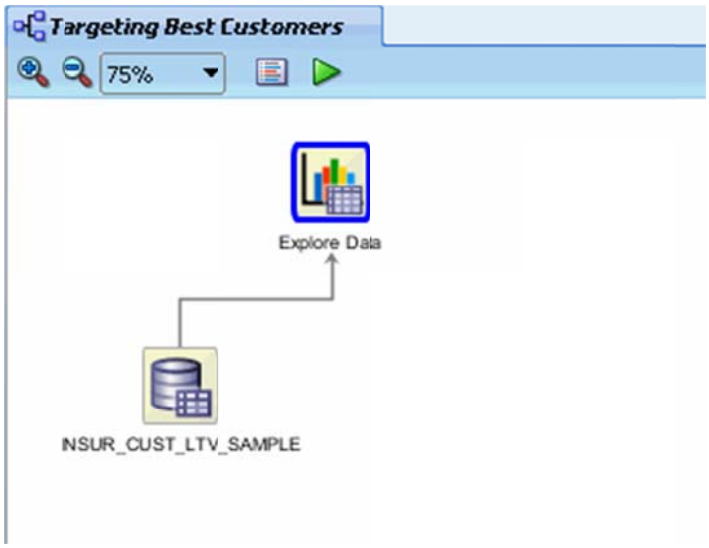
In this case, a data source node must be "linked" to the Explore Data node to enable further exploration of the source data.

2. To link the data source and explore data nodes, use the following instructions:

A. Right-click the data source node (INSUR_CUST_LTV_SAMPLE), select **Connect** from the pop-up menu, and then drag the pointer to the Explore Data node, as shown here:



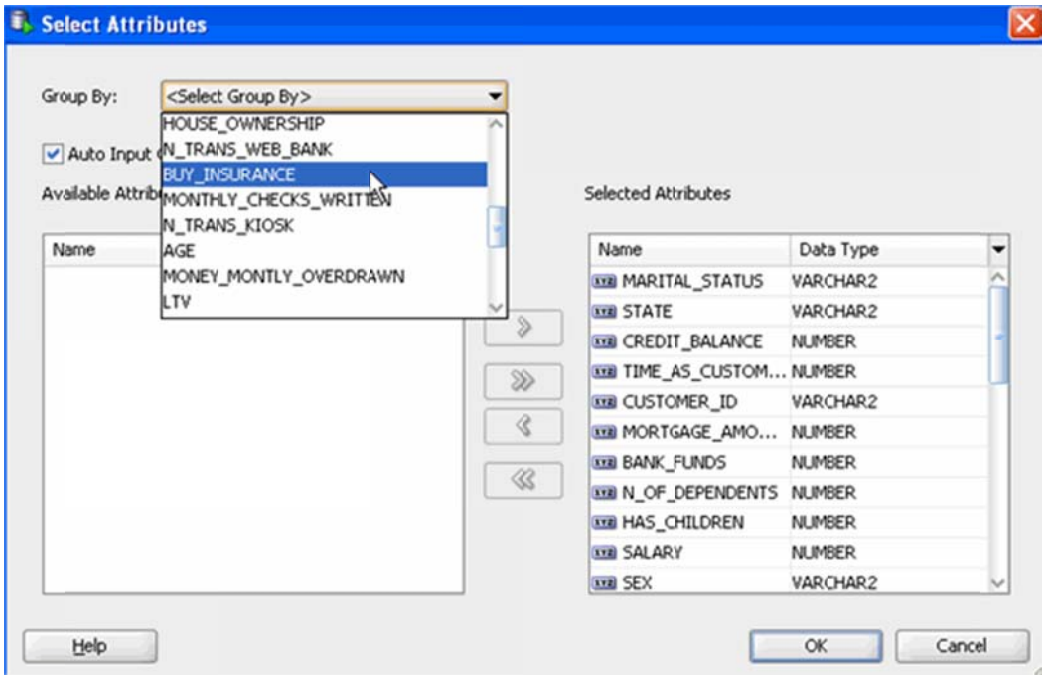
B. Then, click the **Explore Data** node to connect the two nodes. The resulting display looks like this:



3. Next, select a "Group By" attribute for the data source.

A. Double-click the **Explore Data** node to display the Select Attributes window (you may see “Edit Explore Data Node”).

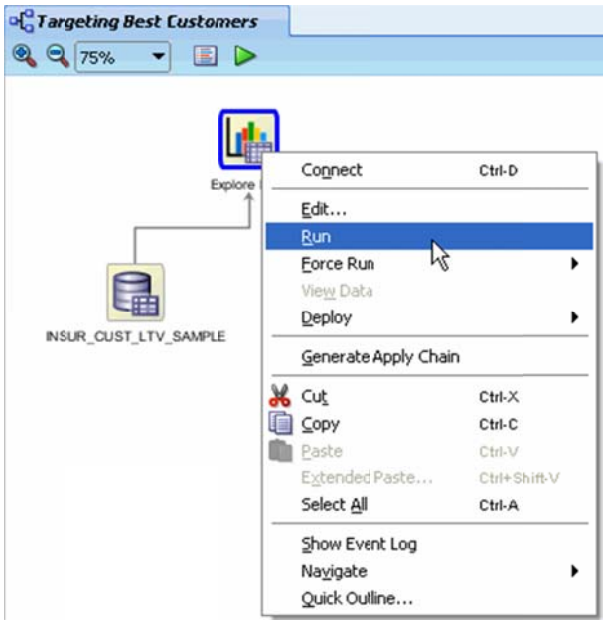
B. In the Group By list, select the **BUY_INSURANCE** attribute, as shown here:



C. Then, click **OK**.

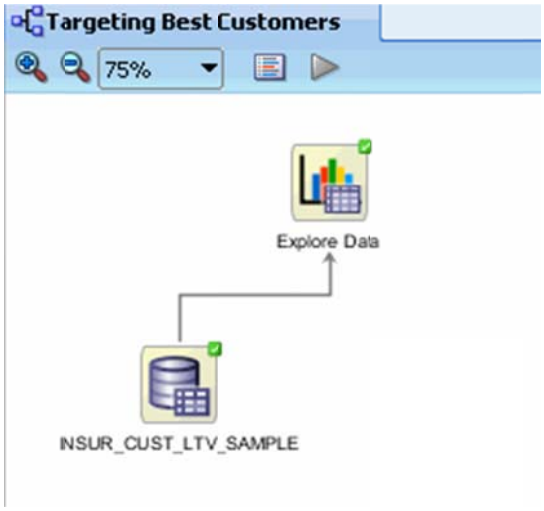
Note: The Select Attributes window also allows you to remove (or re-add) any attributes from the source data.

4 . Next, right-click the Explore Data node and select **Run**.



Result:

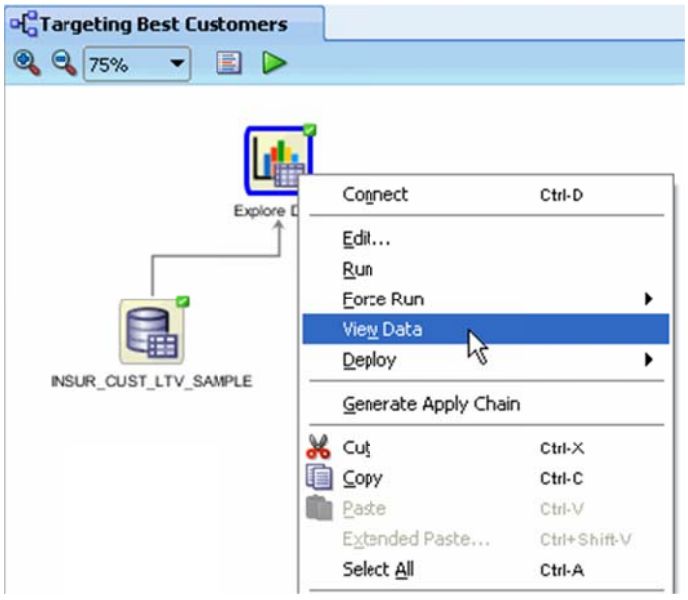
Data Miner displays status information in the Workflow Jobs tab while processing the node.
When the update is complete, the data source and explore data nodes show a green check mark in the borders, like this:



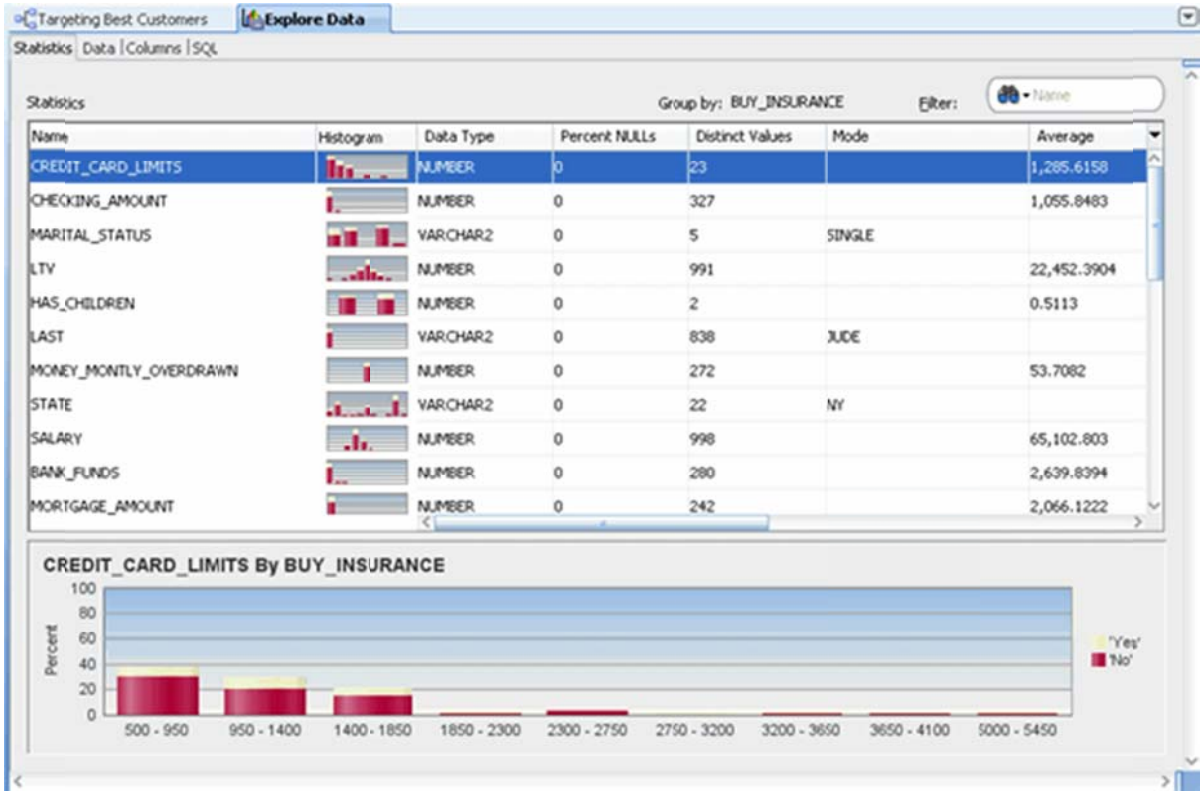
Note: When you run any process from the workflow canvas, the steps that you have specified are executed by the Oracle Data Miner Server.

5. To see results from the Explore Data node, perform the following:

A. Right-click the Explore Data node and select **View Data** from the menu.



Result: A new tab opens for the data profile node, as shown below.



Required: Capture this screen and paste it onto your submission.

Notes:

Data Miner calculates a variety of information about each attribute in the data set, as it relates to the "Group By" attribute that you previously defined, including a Histogram, Distinct Values, Mode, Average, Min and Max value, Standard Deviation, Variance, Skewness, and Kurtosis.

The display enables you to visualize and validate the data, and also to manually inspect the data for patterns or structure.

B. Select any of the attributes in the Name list to display the associated histogram in the bottom window.

C. When you are done examining the source data, dismiss the Expore Data tab by clicking the Close icon (**X**).

Next, you move from a high-level manual analytic exercise to using the power of database data mining.

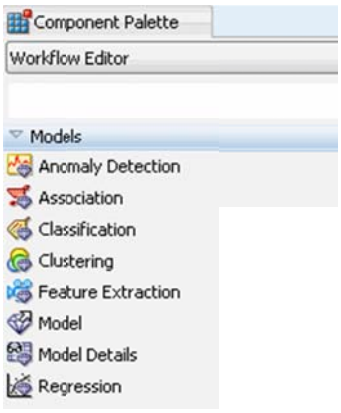
Create Classification Models

As stated in the Overview section of this tutorial, classification models are used to predict individual behavior. In this scenario, you want to predict which customers are most likely to buy insurance. Therefore, you will specify a classification model.

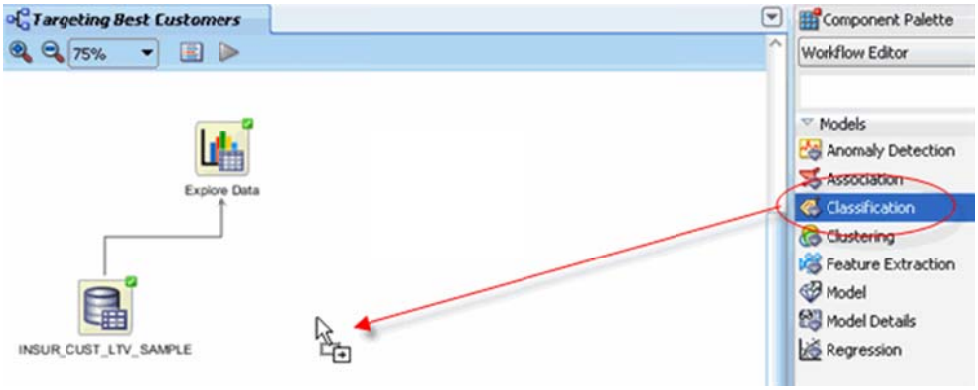
By default, Oracle Data Miner selects all of the supported algorithms for a Classification model. Here, you define a Classification node that uses all algorithms for the model. In the following topic, you will run and examine each of the models.

To create the default Classification models, follow these steps:

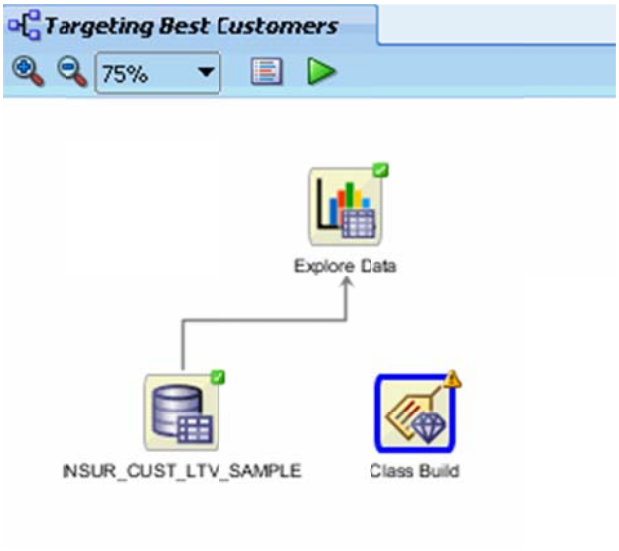
1. A. First, click on **Models** in the Component Palette to display the available list:



B. Then, drag the **Classification** node from the paleltte to the Workflow pane, like this:



C. Drop the node onto the workflow. After a moment, a "Class Build" node appears in the workflow:



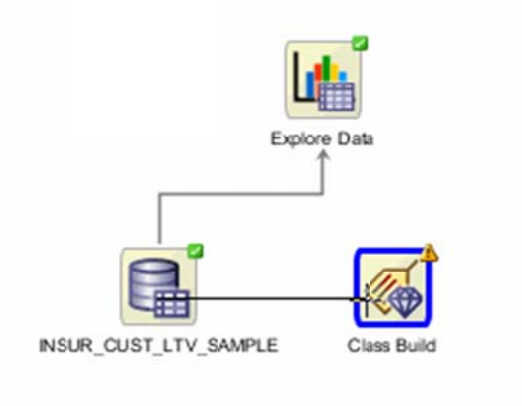
Notes:

As stated previously, a yellow exclamation mark on the border indicates that more information needs to be specified before the node is complete.

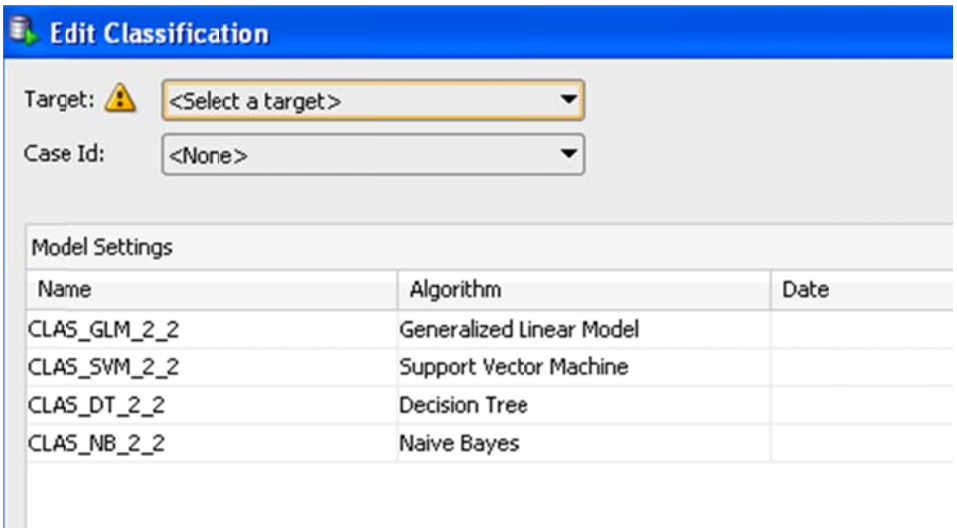
In this case, two actions are required:

1. A link must be created between the source data node and the classification build node.
2. Two attributes should be specified for the classification build process.

2. First, connect the data source node to the classification build node using the same technique described previously.



Result: the Edit Classification window appears (you may see “Edit Classification Build Node” window).



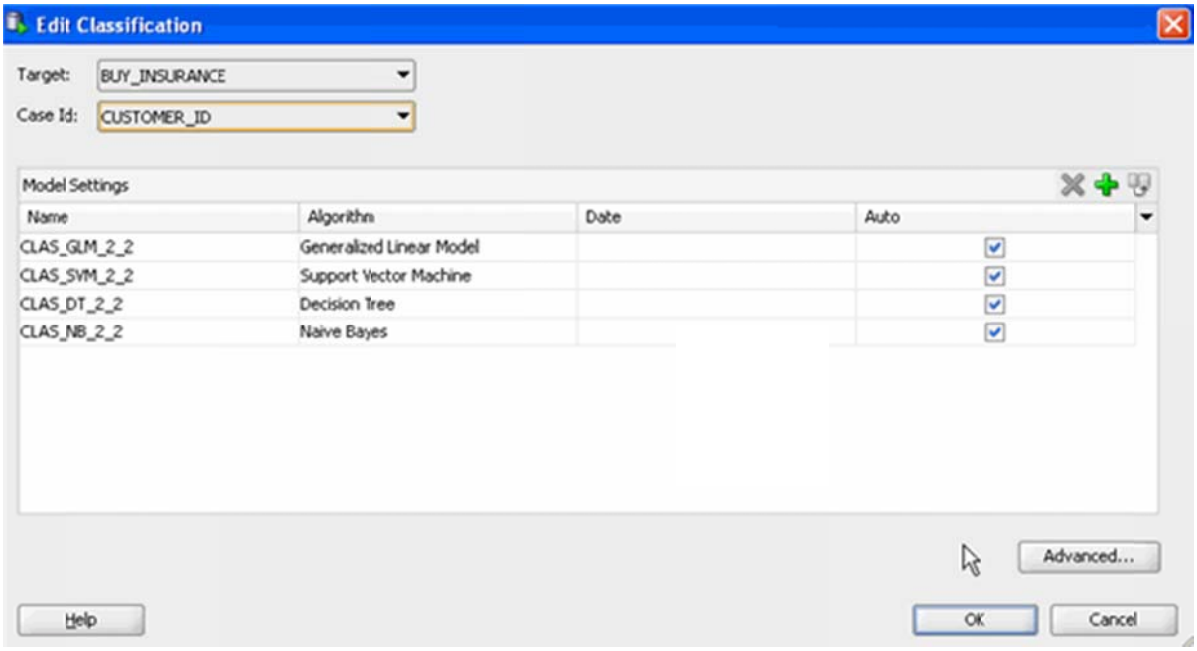
Notes:

Notice that a yellow "!" indicator is displayed next to the Target field. This means that an attribute must be selected for this item.
The names for each model are automatically generated, and yours may differ slightly from those in this example.

3 . In the Edit Classification window:

A. Select **BUY_INSURANCE** as the Target attribute.

B. Select **CUSTOMER_ID** as the Case Id attribute.

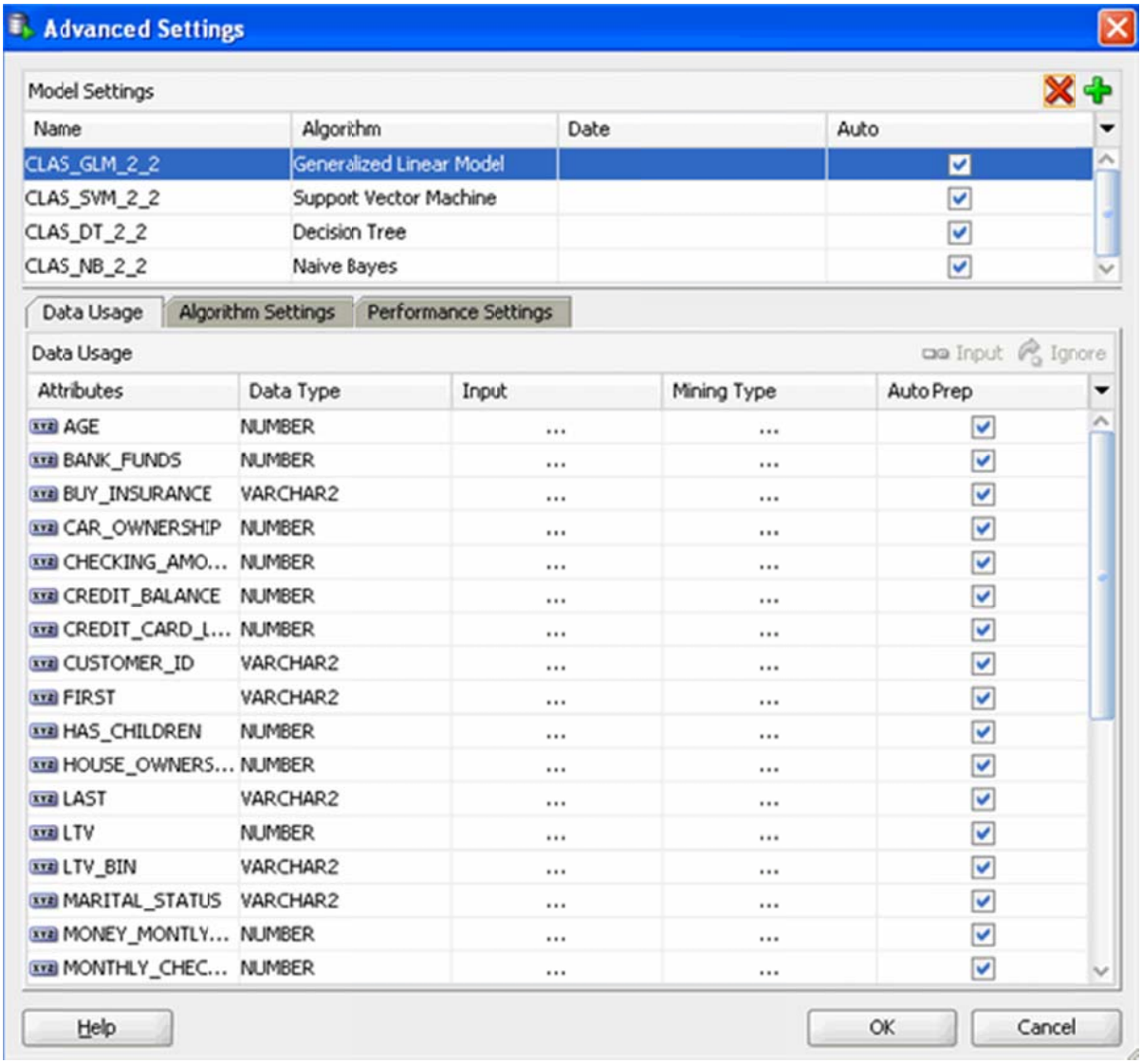


Notes:

Although not required, it is advised that you define a Case Id to uniquely define each record. This helps with model repeatability and is consistent with good data mining practicies.
As stated previously, all four algorithms for Classification modeling are selected by default. They will be automatically run unless you specify otherwise.

4 . Optionally, you can modify specific settings for each of the algorithms by using the **Advanced** button.

A. Click **Advanced** at the bottom of the Edit Classification window to display the Advanced Settings window, as shown here:

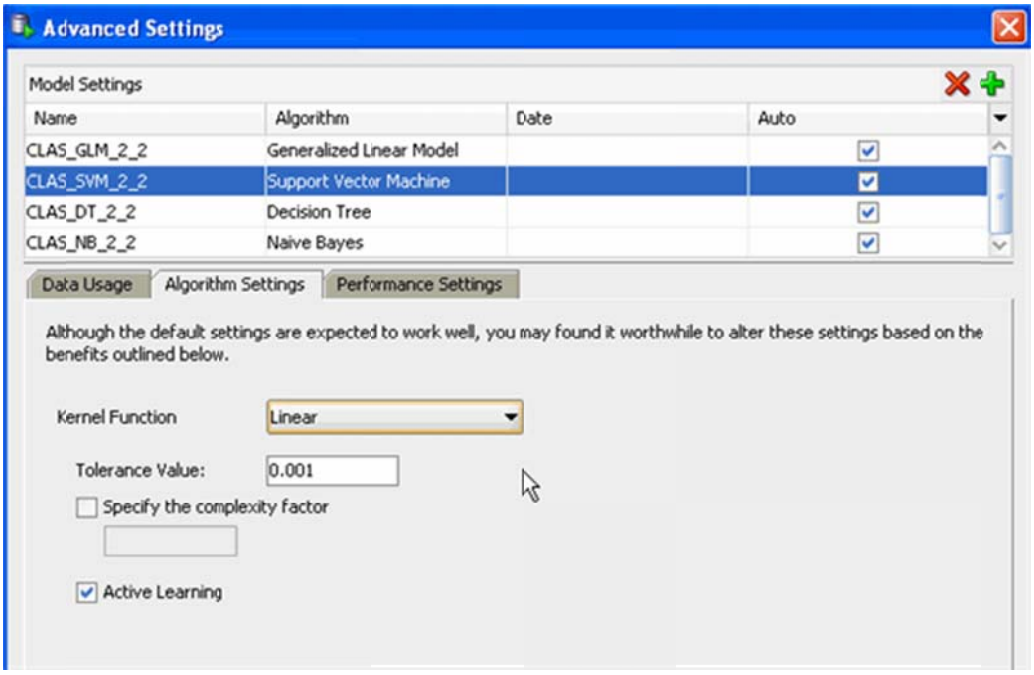


Notes:

The Advanced Settings window enables you to specify data usage, algorithm settings, and performance settings for each of the four classificaiton algorithms.
You can also de-select (and re-select) any algorithm from this window.

B. Select the **Support Vector Machine** algorithm and click the **Algorithm Settings** tab.

C. Then, In the Kernel Function option, select **Linear**, as shown here:



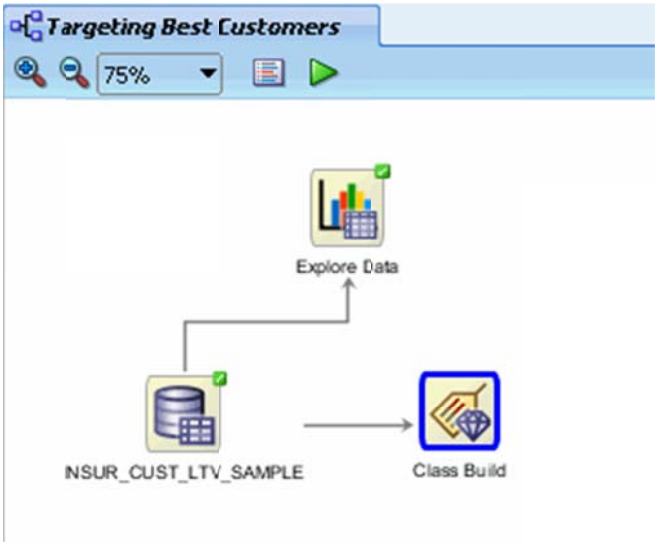
Note: We want to change the value of this Support Vector Machine (SVM) algorithm setting from the system determined value to Linear in order to make the model results easier to interpret.

D. Feel free to view any of the tabs for each algorithm, however do not modify any of the other default settings.

E. When you are done browsing, click **OK** to save the SVM algorithm setting and close the Advanced Settings window.

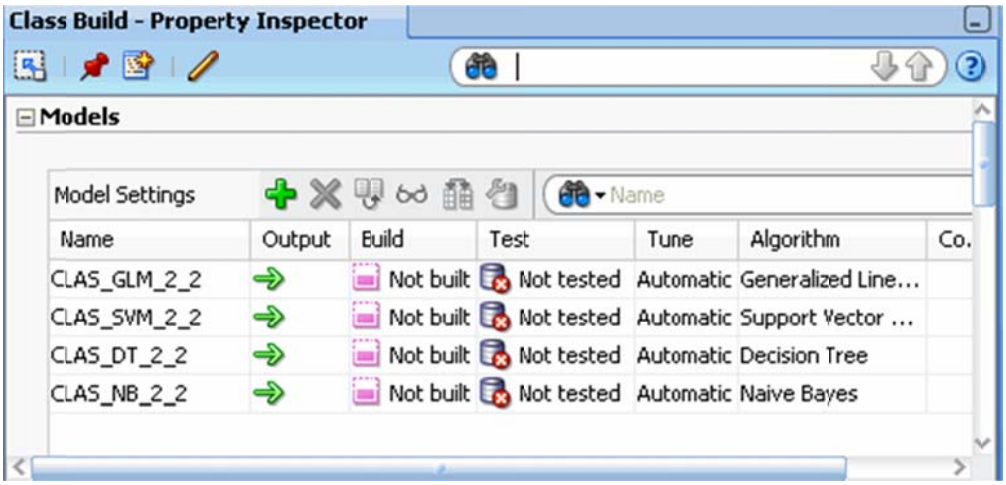
5 . Finally, click **OK** in the Edit Classification window to save your changes.

Result: The classification build node is ready to run.

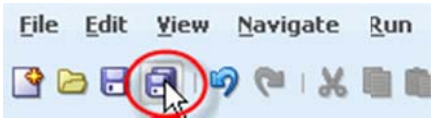


Note: In the **Models** tab of the Properties Inspector, you can see the current status for each of the selected

algorithms, as shown below:



6 . Save the workflow by clicking the **Save All** icon in main toolbar.

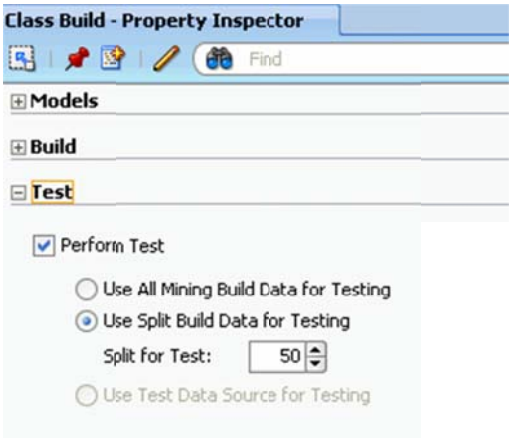


Build the Models

In this topic, you build the selected models against the source data. This operation is also called “training” a model, and the model is said to “learn” from the training data.

A common data mining practice is to build (or train) your model against part of the source data, and then to test the model against the remaining portion of your data. By default, Oracle Data Miner this approach.

Before building the models, select Class Build node and choose the **Test** tab in the Property Inspector. Then, change the split to **50**, as shown here:



With this setting, Oracle Data Miner will split the build data in a 50/50 fashion.

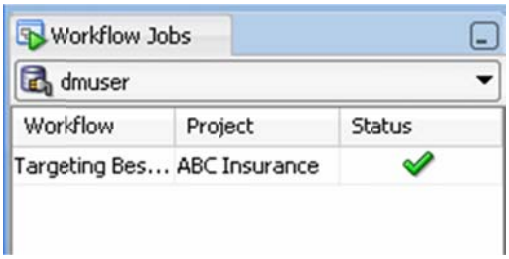
To build the models, run the Classification Build node.

1. Right-click the classification build node and select **Run** from the pop-up menu.

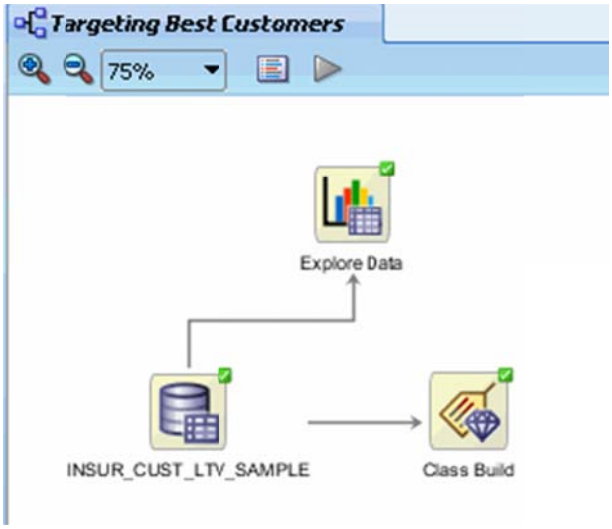
   Notes:

   When the node runs it builds and tests all of the models that are defined in the node.
   As before, a green gear icon appears on the node border to indicate a server process is running, and the Workflow Jobs tab shows the status of the build.

   When the build is complete, the status column displays a green check mark.

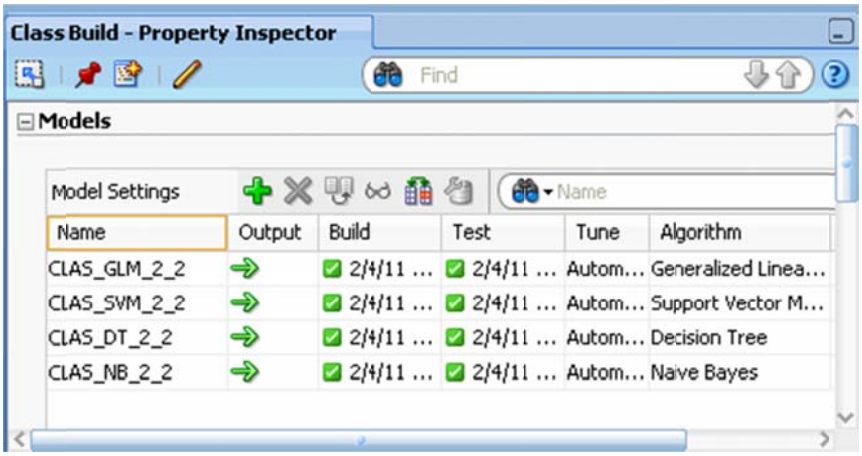


   In the workflow pane, the border of the build node changes from a green gear turning to a green check mark, like this:



2. Once the build process is complete, you can view several pieces of information about the build using the property inspectory.

   A. Select the classification build node in the workflow, and then choose the **Models** tab in the Property Inspector.



Required: Capture this screen and paste it onto your submission

Notes:

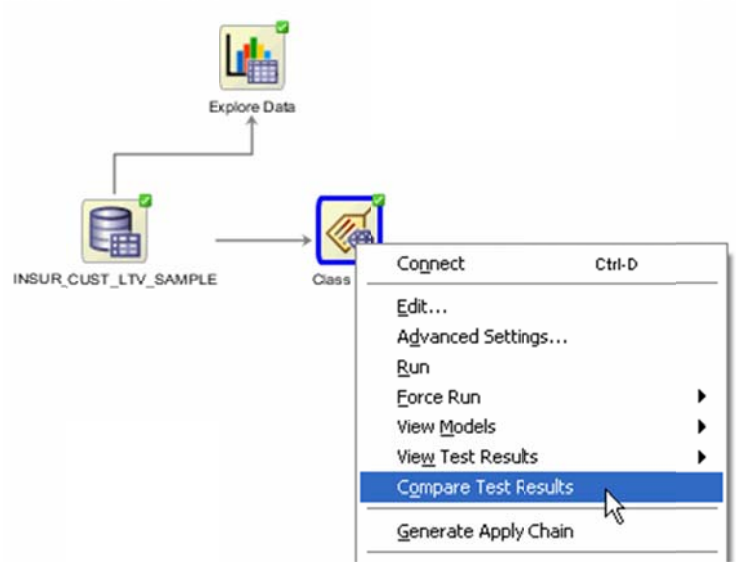
All four models have been succesfully built.
The models all have the same target (BUY_INSURANCE) but use different algorithms.
The source data is automatically divided into test data and build data.

Compare the Models

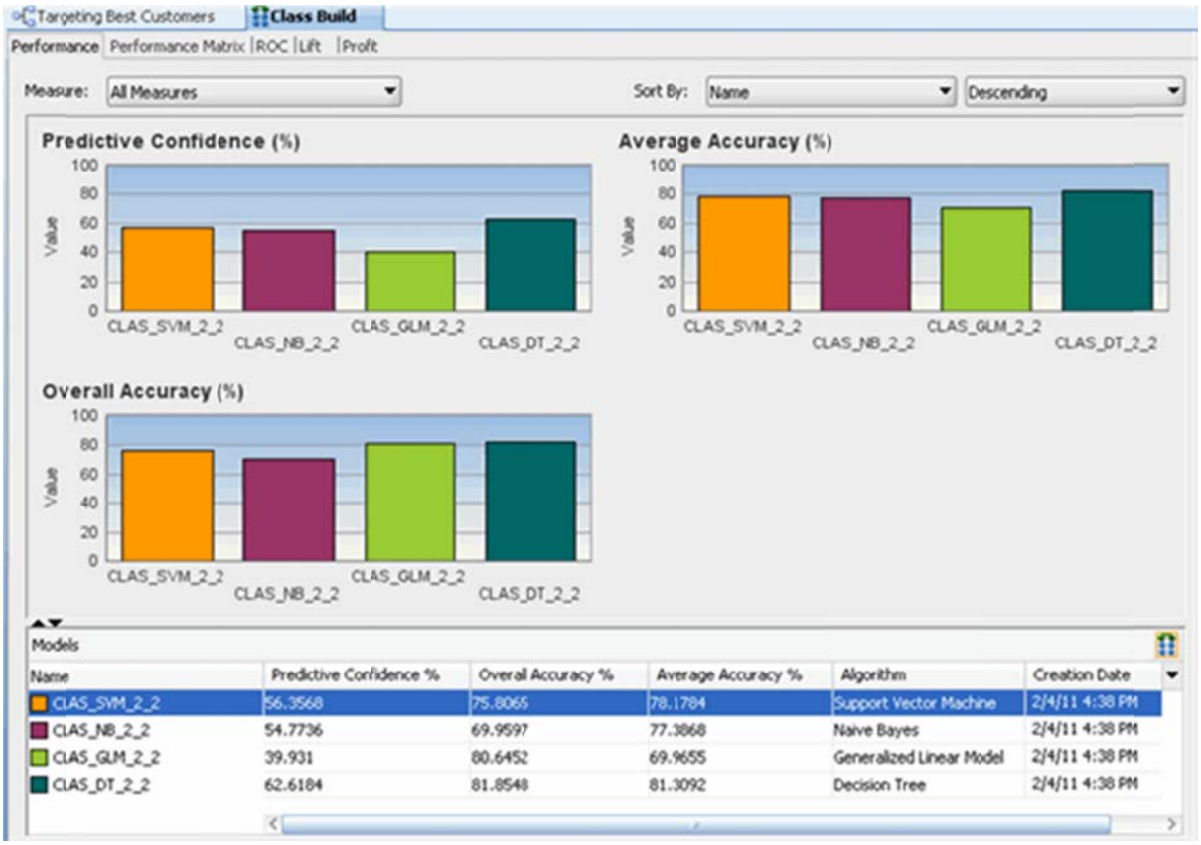
After you build/train the selected models, you can view and evaluate the results for all of the models in a comparative format. Here, you compare the relative results of all four classification models.

Follow these steps:

1. Right-click the classification build node and select **Compare Test Results** from the menu.



Results: A Class Build display tab opens, showing a graphical comparison of the four models, as shown here:



Required: Capture this screen and paste it onto your submission

Notes:

Since the sample data set is very small, the numbers you get may differ slightly from those shown in the tutorial example. In addition, the histogram colors that you see may be different then those shown in this example.

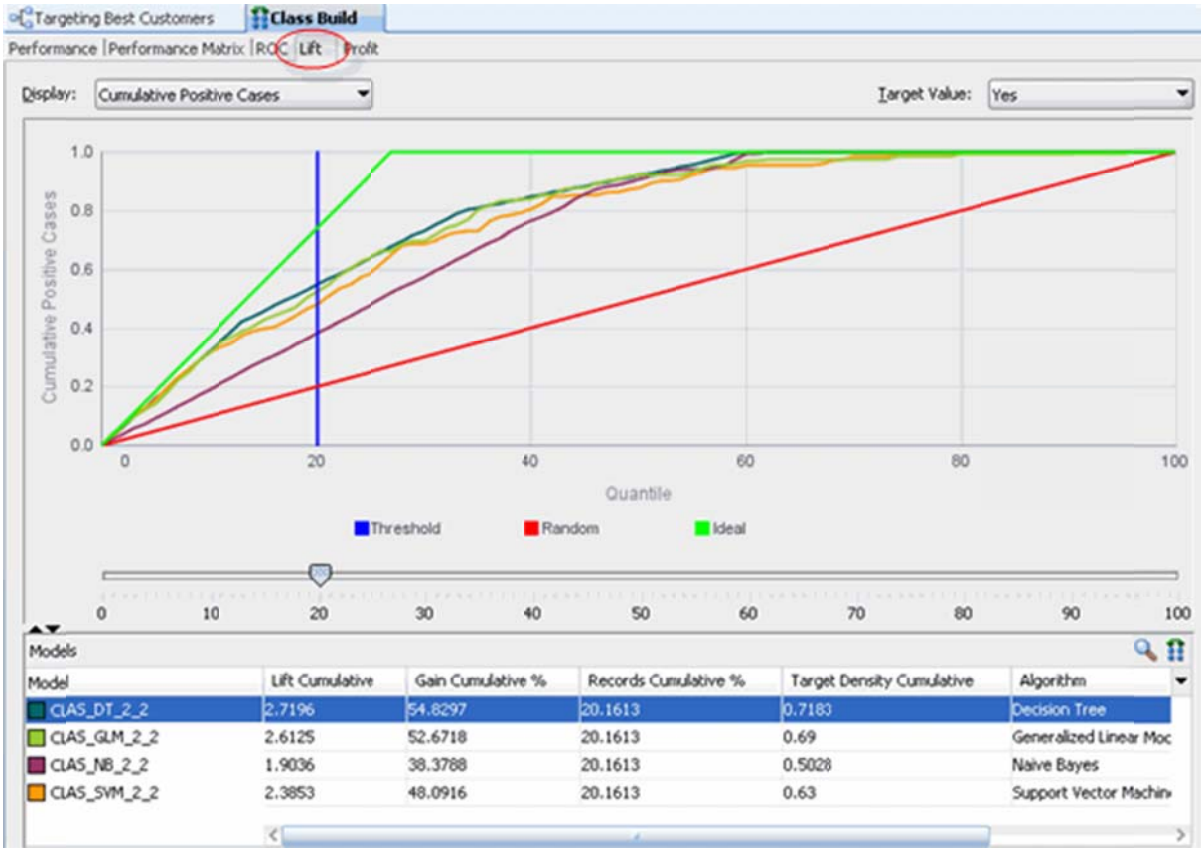
The comparison results include five tabs: Performance, Performance matrix, ROC, Lift, and Profit.

The Performance tab provides numeric and graphical information for each model on Predictive Confidence, Average Accuracy, and Overall Accuracy.

The Performance tab seems to indicate that:

- The Decision Tree (DT) is providing the highest Predictive Confidence.
- Both the DT and Generalized Linear Model (GLM) are providing the highest Overall Accuracy results.

2 . Select the **Lift** tab.



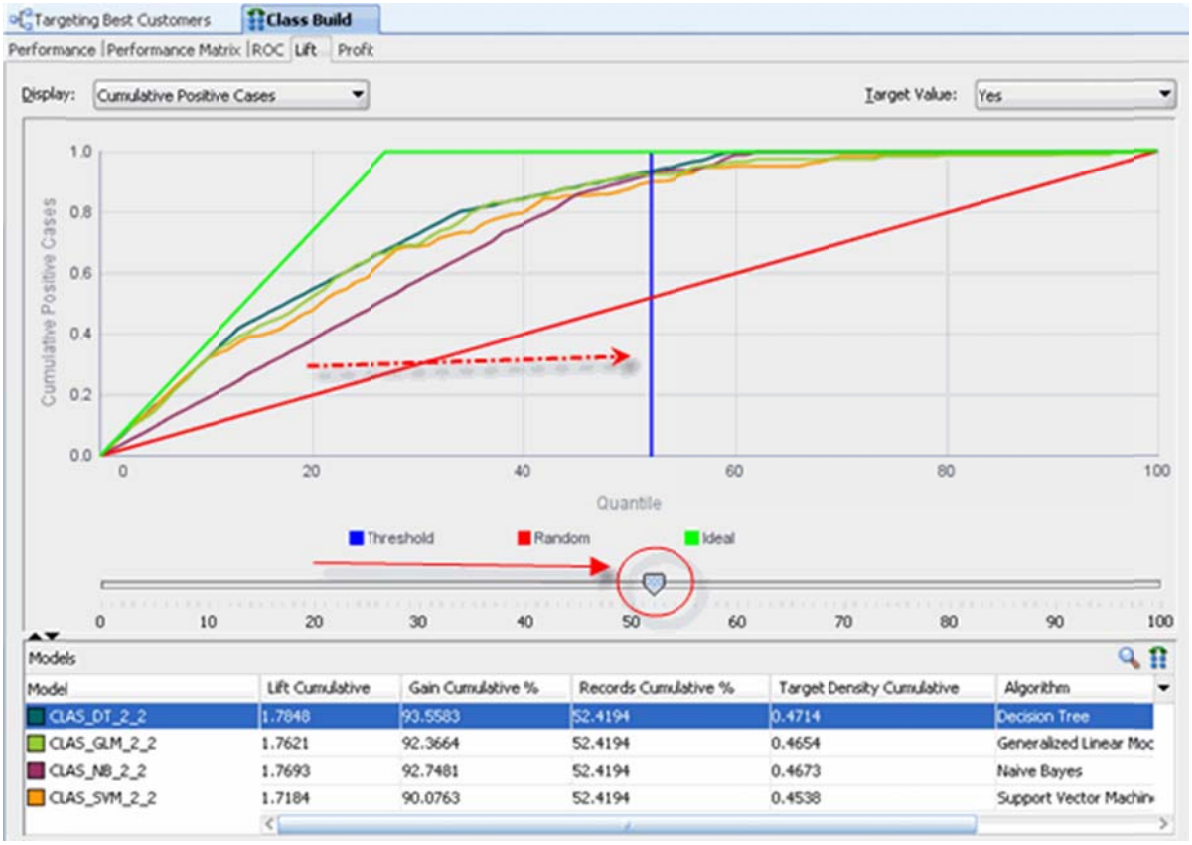
Notes:

The Lift tab provides a graphical presentation showing lift for each model, a red line for the random model, and a vertical blue line for threshold.

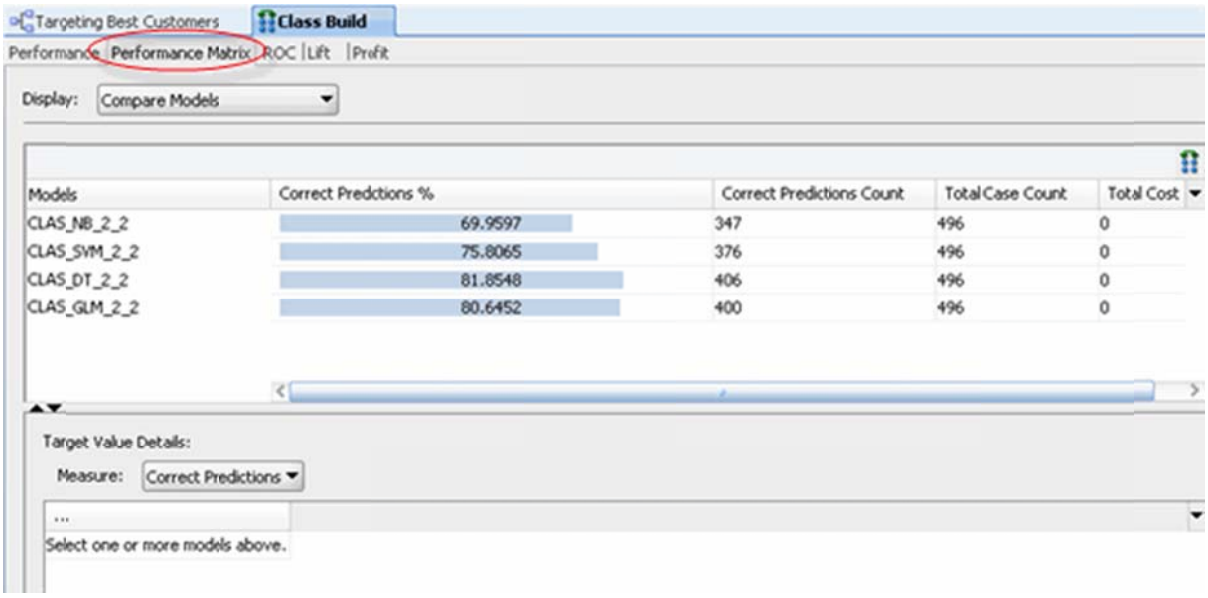
- Lift is a different type of model test. It is a measure of how "fast" the model finds the actual positive target values.
- The Lift viewer compares lift results for the given target value in each model.
- The Lift viewer displays Cumulative Positive Cases and Cumulative Lift.

Using this example:

- At the 20th quantile, the DT model provides the greatest Cumulative Lift, just above the GLM model (see above).
- You can move the Quanitile measure point line along the X axis of the graph by using the slider tool. The data in the Models pane at the bottom updates automatically as you move the slider left or right.
- As you move up the quantile range, the Cumulative Lift of the GLM and Support Vector Machine (SVM) models fluxuate near or below that of the DT model, and the Naive Bayes (NB) model shows more of a linear increase, but still below that of the DT model.
- At just above the 52nd quantile (see below), the NB and GLM models provide almost the same Cumulative Lift as the DT model, but none of the other models surpass the lift results of the DT model.



3 . Next, select the **Performance Matrix** tab.



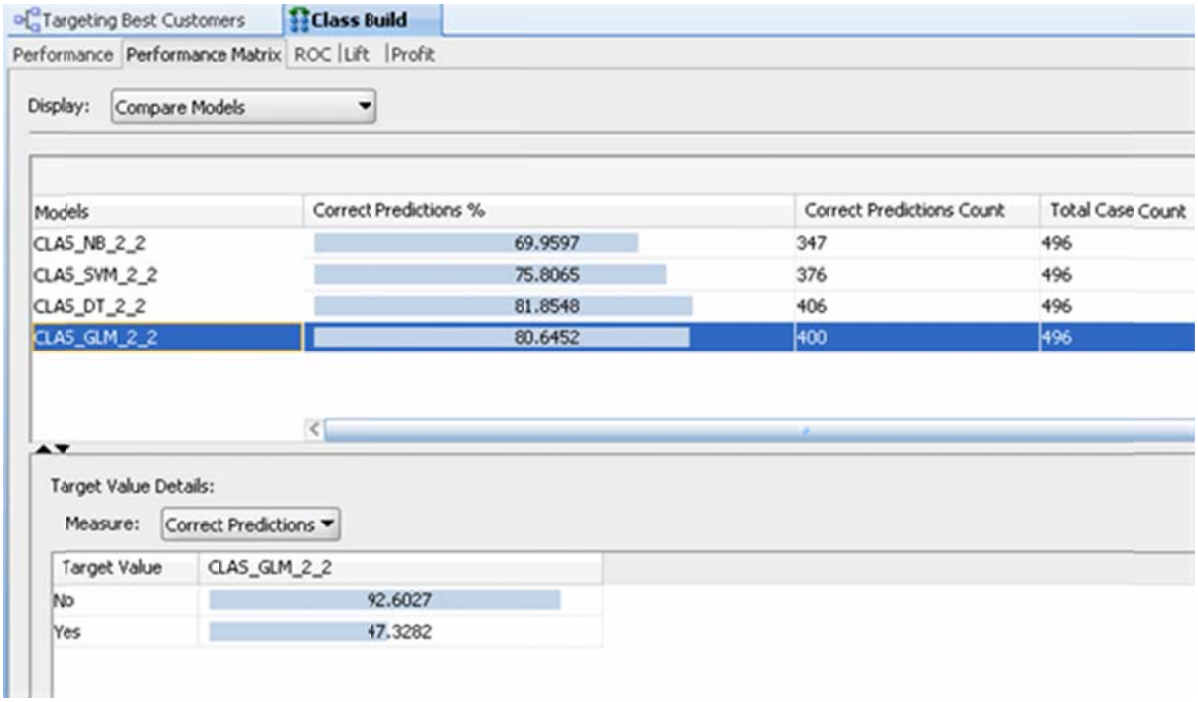
Required: Capture this screen and paste it onto your submission

Note: The Performance Matrix shows that the GLM and DT models have a higher Correct Prediction percentage than

the other models, at over 80% each.

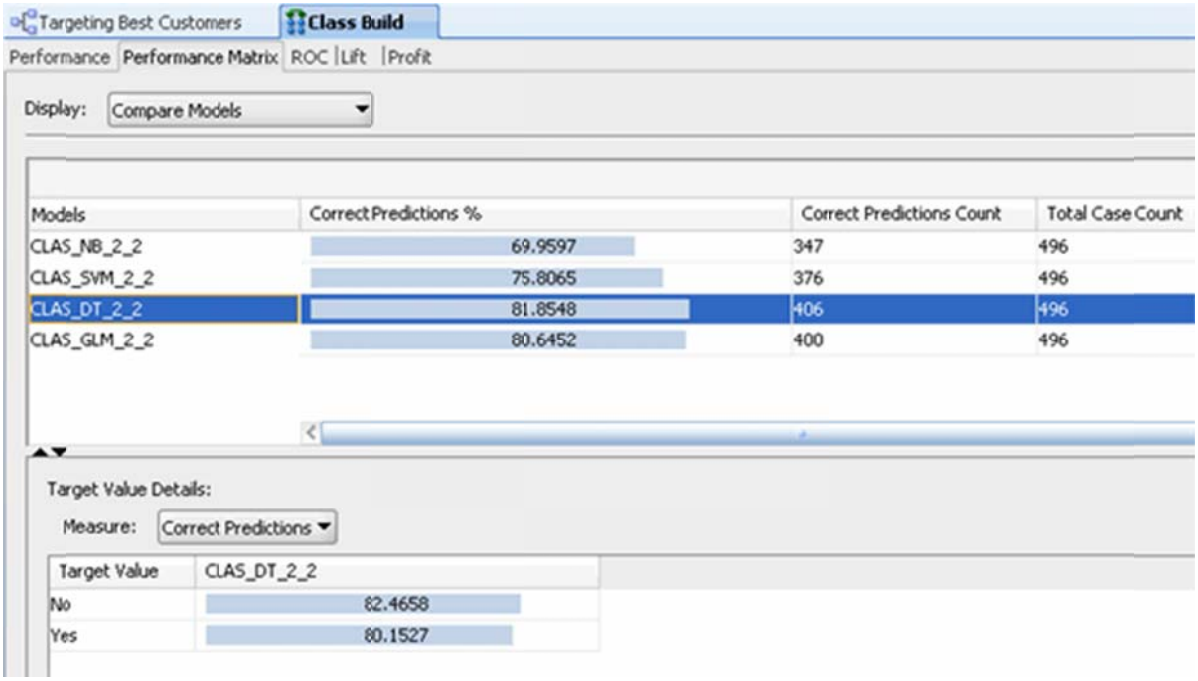
4 . Compare the details for the GLM and DT models.

First, select the **GLM** model to view the Target Value Details for this model. Recall that the "Target Value" for each of the models is the BUY_INSURANCE attribute.



Note: The GLM model indicates a 92.6% correct prediction outcome for customers that don't buy insurance, but only a 47.3% correct prediction outcome for customers that do buy insurance.

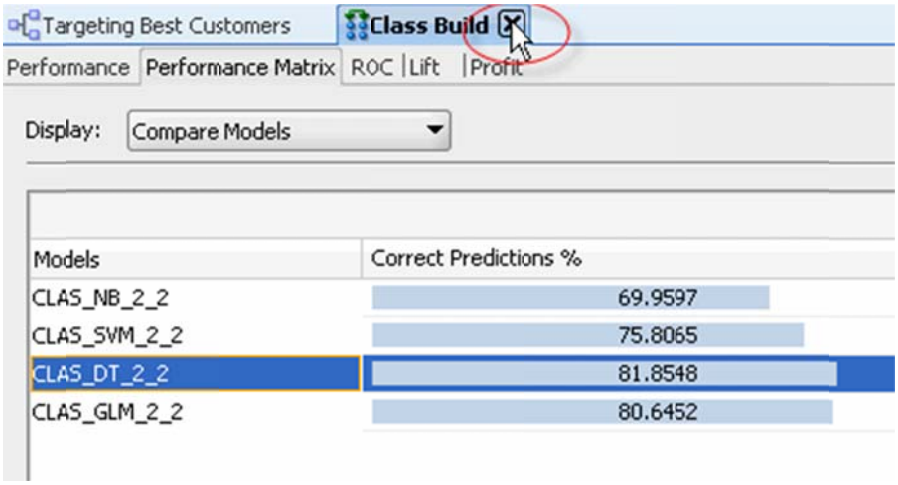
Next, select the DT model.



Notes: The DT model indicates an 82.5% correct prediction outcome for customers that don't buy insurance, and an 80.2% correct prediction outcome for customers that do buy insurance.

5 . After considering the initial analysis, you decide to investigate the Decision Tree model more closely.

First, dismiss the **Class Build - Compare Models** tab, as shown here:



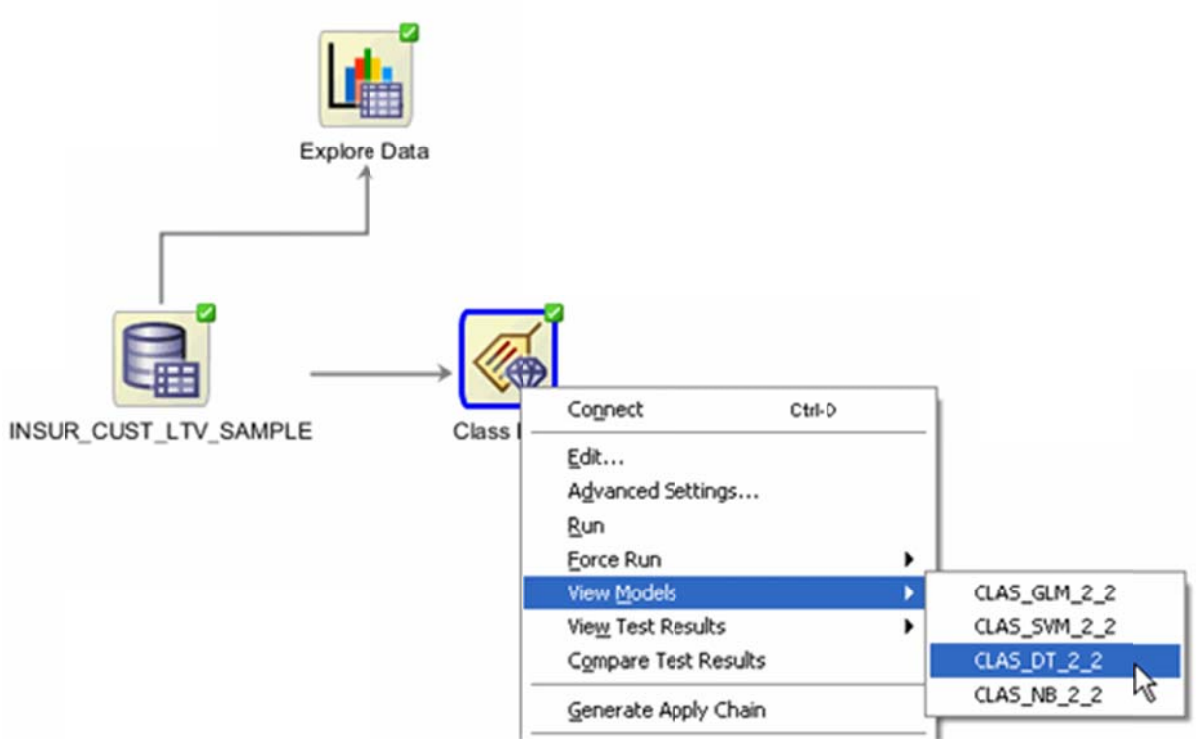
Select and Examine a Specific Model

Using the analysis performed in the past topic, the Decision Tree model is selected for further analysis.

Follow these steps to examine the Decision Tree model.

1 . Back in the workflow canvas, right-click the Class Build node again, and select **View Models** > **CLAS_DT_2_2**(Note:

The exact name of your Decision Tree model may be different).



Result: A window opens that displays a graphical presentation of the Decision Tree.

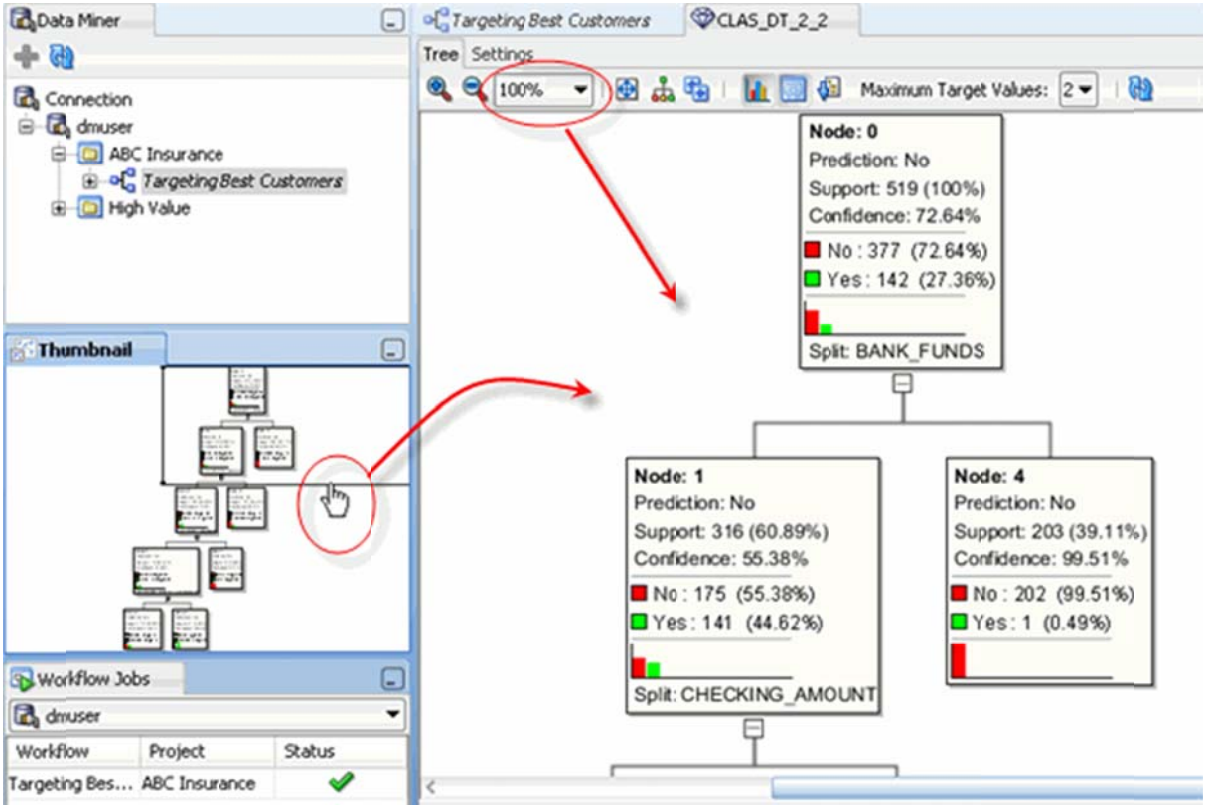
2 . The interface provides several methods of viewing navigation:

The Thumbnail tab provides a high level view of the entire tree. For example, the Thumbnail tab shows that this tree contains five levels, although you view fewer of the nodes in the primary display window.

You can move the viewer box around within the Thumbnail tab to dynamically locate your view in the primary window. You can also use the scroll bars in the primary display window to select a different location within the decision tree display.

Finally, you can change the viewer percentage zoom in the primary display window to increase or descrease the size of viewable content.

For example, set the the primary viewer window for the decision tree to **100%** zoom. This provides a more narrow view than 75%, but the content within the view is somewhat larger.



3 . First, navigate to and select **Node 2** and click on it to select it.

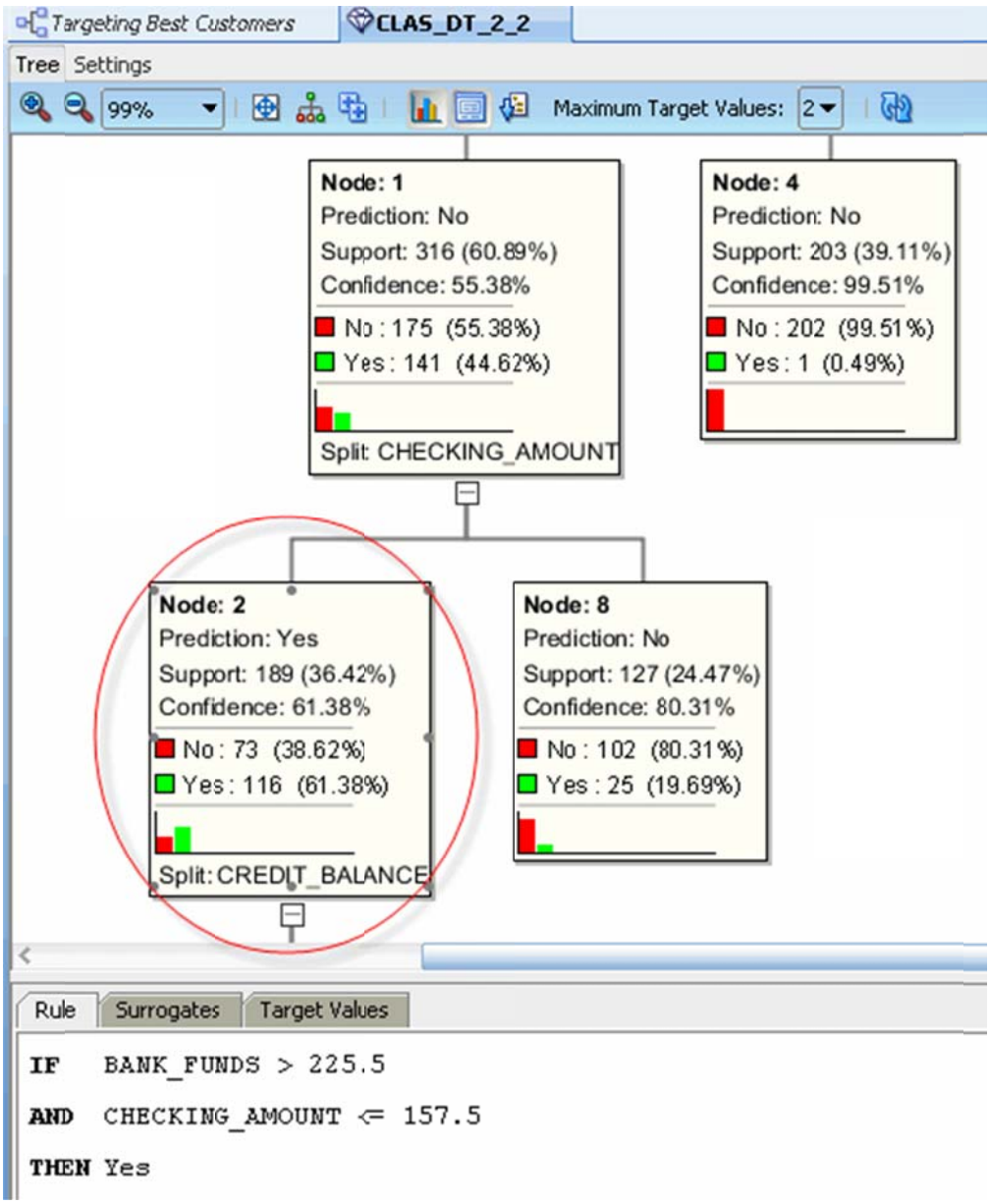
Notes:

At each level within the decision tree, an IF/THEN statement that describes a rule is displayed. As each additional level is added to the tree, another condition is added to the IF/THEN statement.

For each node in the tree, summary information about the particular node is shown in the box.

In addition, the IF/THEN statement rule appears in the Rule tab, as shown below, when you select a particular node.

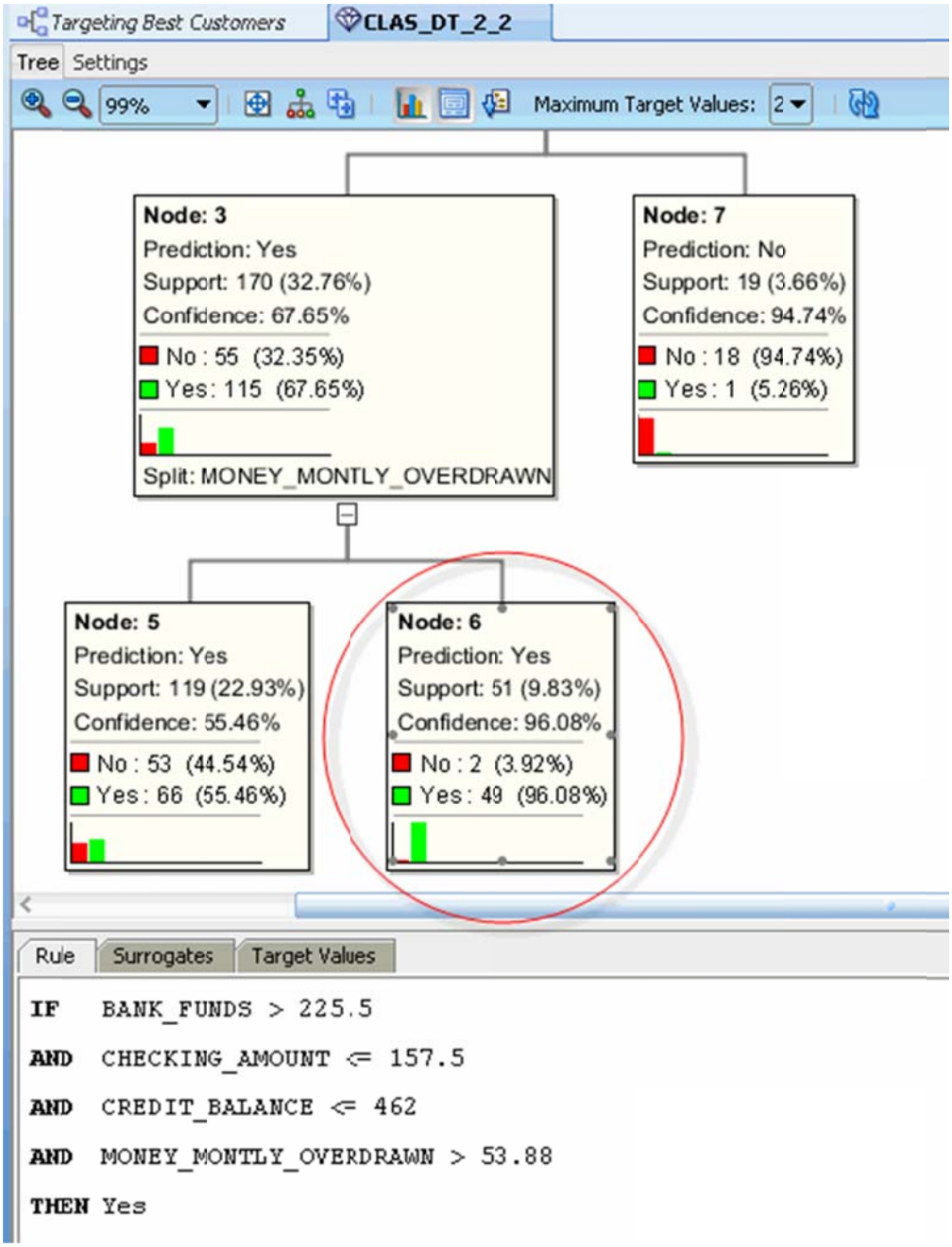
Commonly, a decision tree model would show a much larger set of levels and also nodes within each level in the decision tree. However, the data set used for this lesson is significantly smaller than a normal data mining set, and therefore the decision tree is also small.



Notes:

At this level, we see that the first split is based on the BANK_FUNDS attribute, and the second split is based on the CHECKING_AMOUNT attribute.
Node 2 indicates that if BANK_FUNDS are greater than 225.5, and CHECKING_AMOUNT is less than or equal to 157.5, then there is a 61.38% chance that the customer will buy insurance.

4 . Next, select **Node 6**, at the bottom level in the tree.

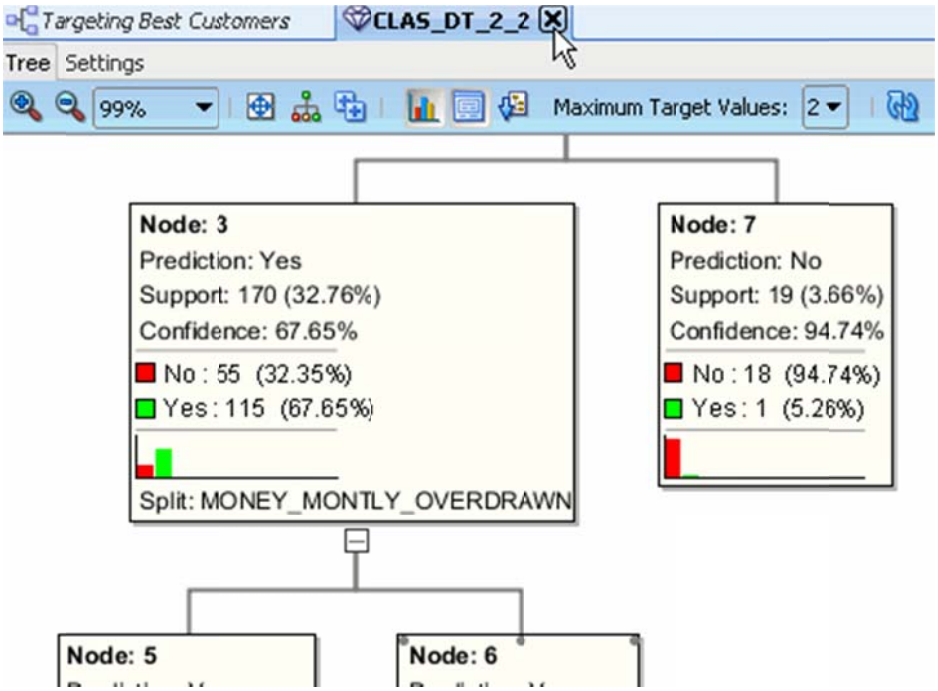


Required: Capture this screen and paste it onto your submission

Notes:

At this bottom level in the tree, a final split is added for the MONEY_MONTHLY_OVERDRAWN attribute. This node indicates that if BANK_FUNDS are greater than 225.5, and CHECKING_AMOUNT is less than or equal to 157.5, and CREDIT_BALANCE is less than or equal to 462, and MONEY_MONTHLY_OVERDRAWN is greater than 53.88, then there is a 96% chance that the customer will buy insurance.

5 . Dismiss the Decision Tree display tab as shown here:



Apply the Model

In this topic, you apply the Decision Tree model and then create a table to display the results. You "apply" a model in order to make predictions - in this case to predict which customers are likely to buy insurance.

To apply a model, you perform the following steps:

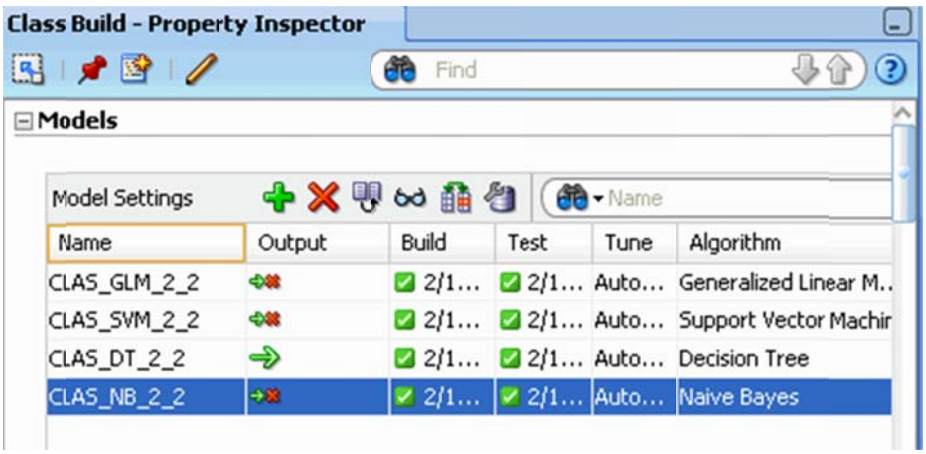
1. First, specify the desired model (or models) in the Class Build node.
2. Then, you add a new Data Source node to the workflow. (This node will serve as the "Apply" data.)
3. Next, add an Apply node to the workflow.
4. Finally, you link both the Class Build node and the new Data Source node to the Apply node.

Follow these steps to apply the model and display the results:

1. In the workflow, select the Class Build node. Then, using the Models tab of the Property Inspector, deselect all of the models except for the DT model.

To deselect a model, click the large green arrow in the model's **Output** column. This action adds a small red "**x**" to the column, indicating that the model will not be used in the next build.

When you finish, the Models tab of the Property Inspector should look like this:

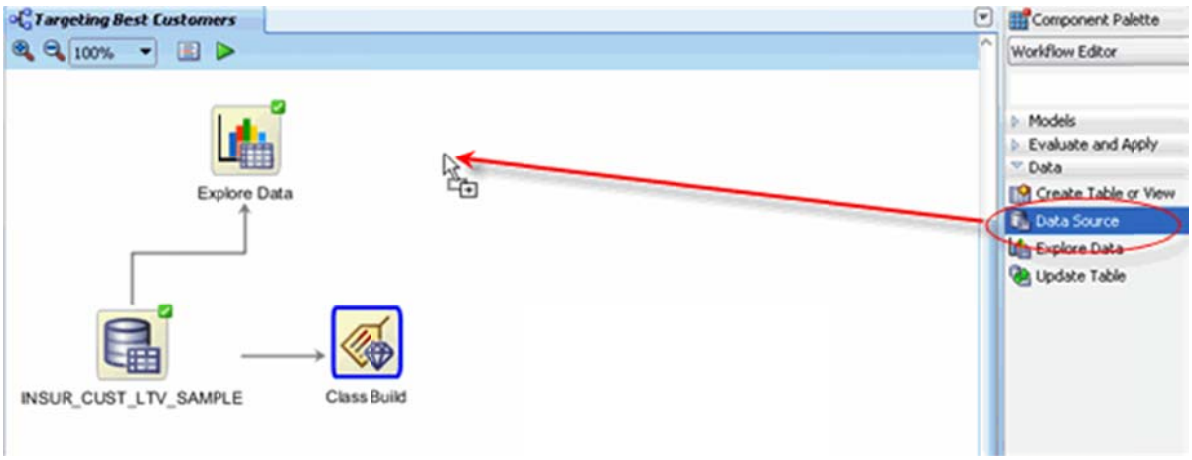


Note: Now, only the DT model will be passed to subsequent nodes.

2. Next, create a new Data Source node in the workflow.

Note: Even though we are using the same table as the "Apply" data source, you must still add a second data source node to the workflow.

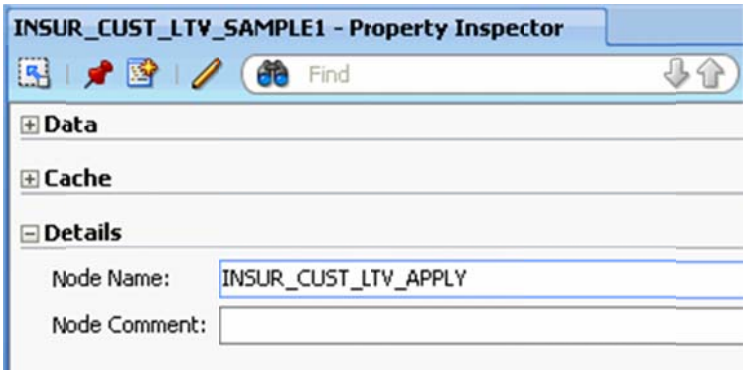
A. From the Data list in the Component Palette, drag and drop the **Data Source** node to the workflow canvas, like this:



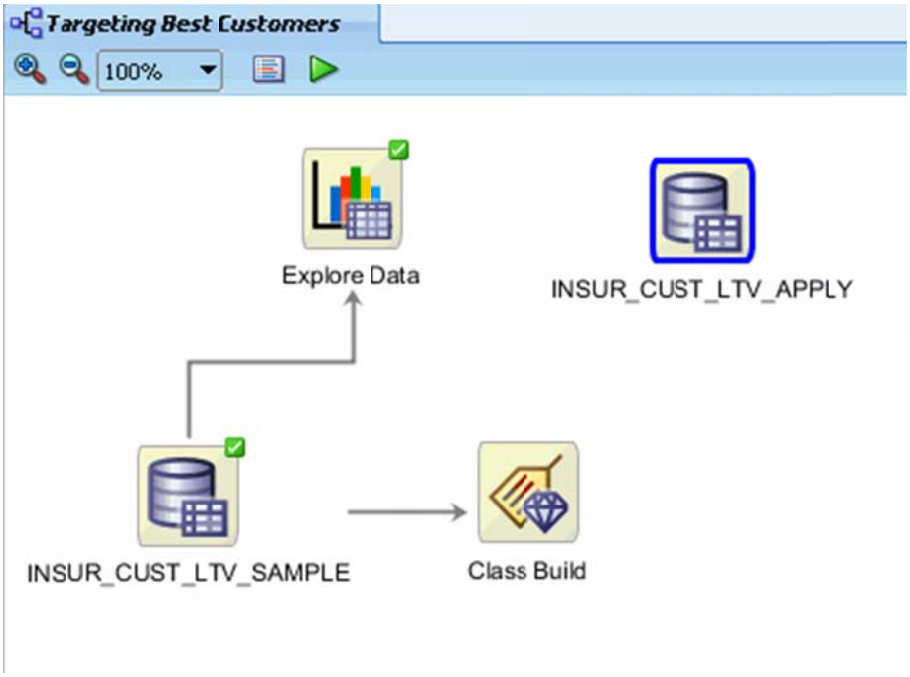
B. In the Define Data Source wizard, select the **INSUR_CUST_LTV_SAMPLE** table, and then click **FINISH**.

Result: A new data souce node appears on the workflow canvas.

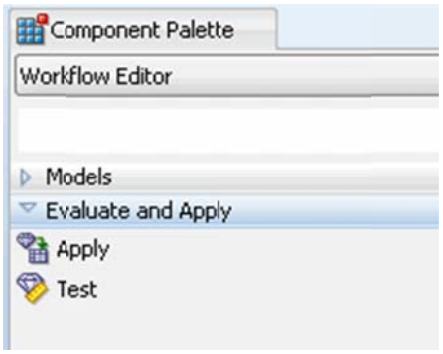
C. Select the new data source node, and using the **Details** tab of the Property Inspector, change the Node Name to **INSUR_CUST_LTV_APPLY**, like this:



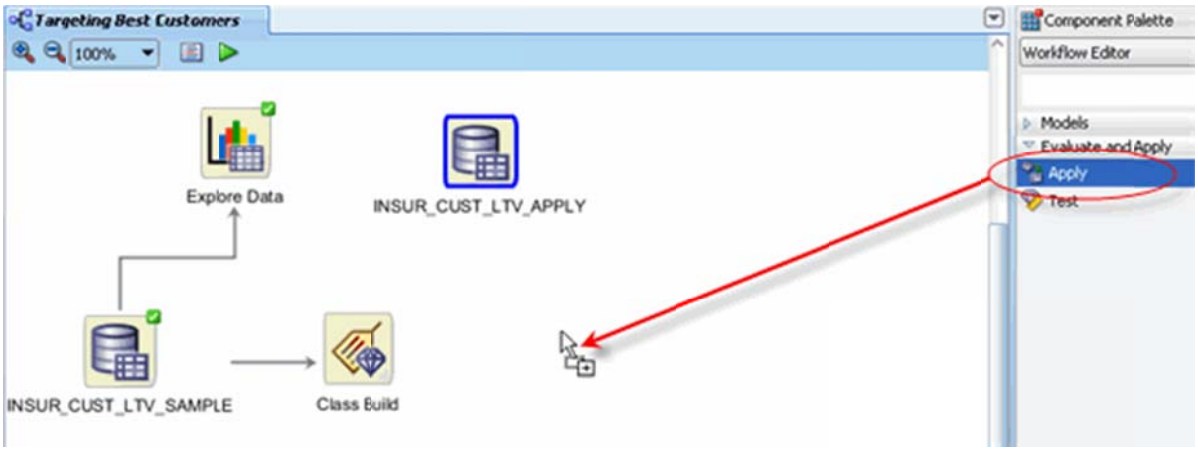
Result: The new table name is reflected in the workflow.



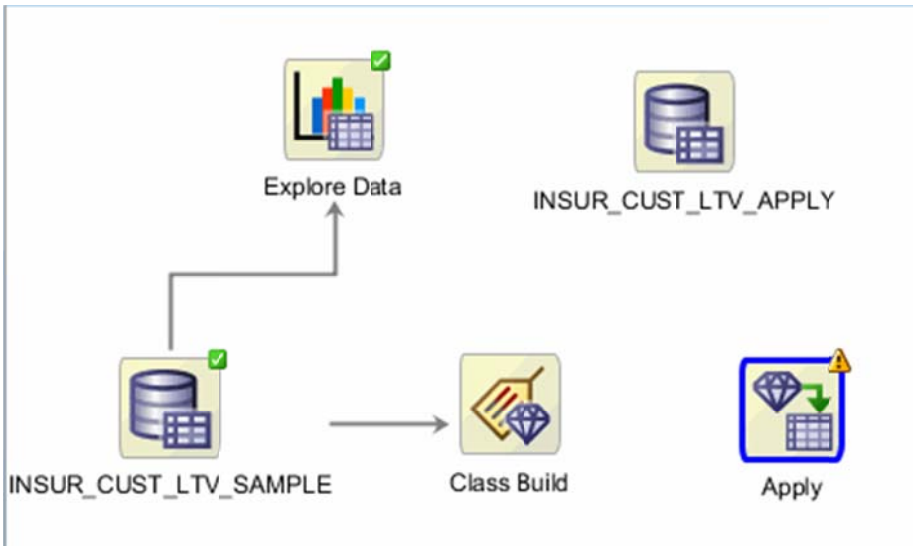
3 . Next, open the **Evaluate and Apply** list in the Components Palette, like this:



4 . A. Drag and drop the **Apply** node to the workflow canvas, like this:

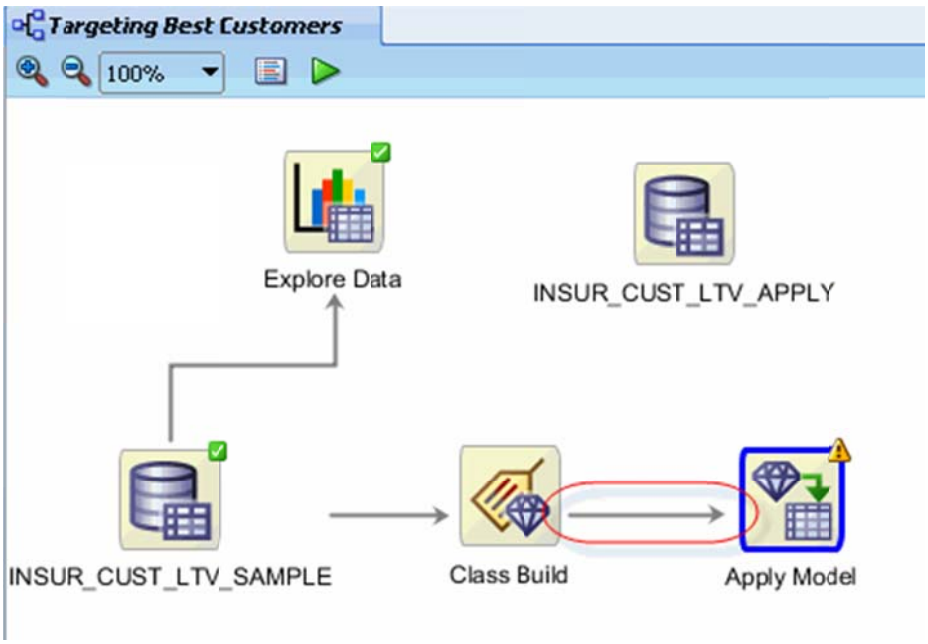


Result: An Apply node is added to the workflow with a yellow exclamation mark in its border. This, of course, indicates that more information is required before this node may be run.

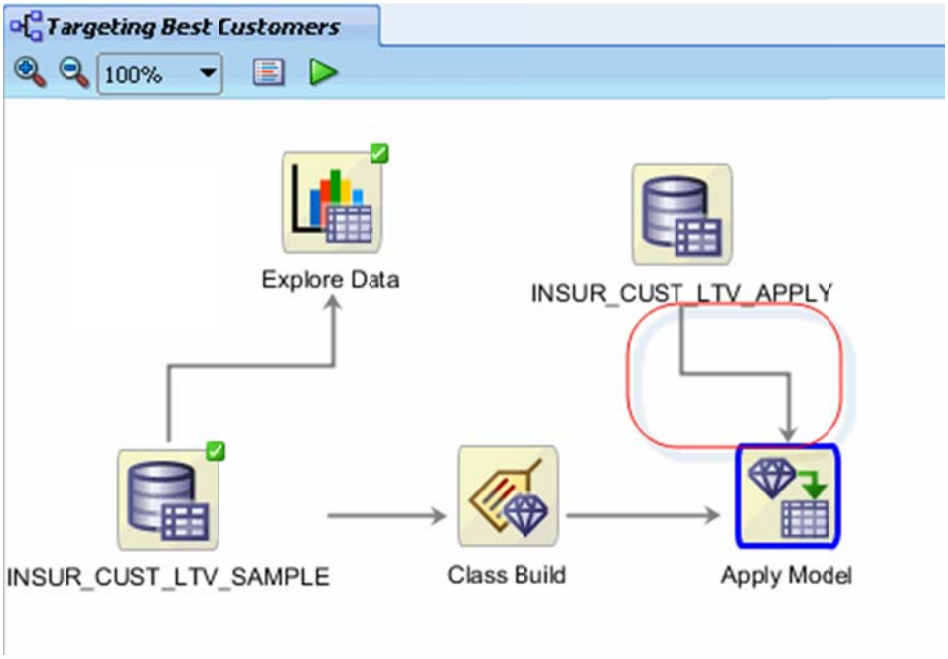


B. Using the Details tab of the Property Inspectory, rename the Apply node to **Apply Model**.

5 . Using the techniques described previously, connect the **Class Build** node to the **Apply Model** node, like this:



6 . Then, connect the **CUST_INSUR_LTV_APPLY** node to the **Apply Model** node:



Notes:

The yellow exclamation mark disappears from the Apply node border once the second link is completed. This indicates that the node is ready to be run.

7 . Before you execute the apply model node, consider the resulting output. By default, an apply node creates two

columns of information for each customer:

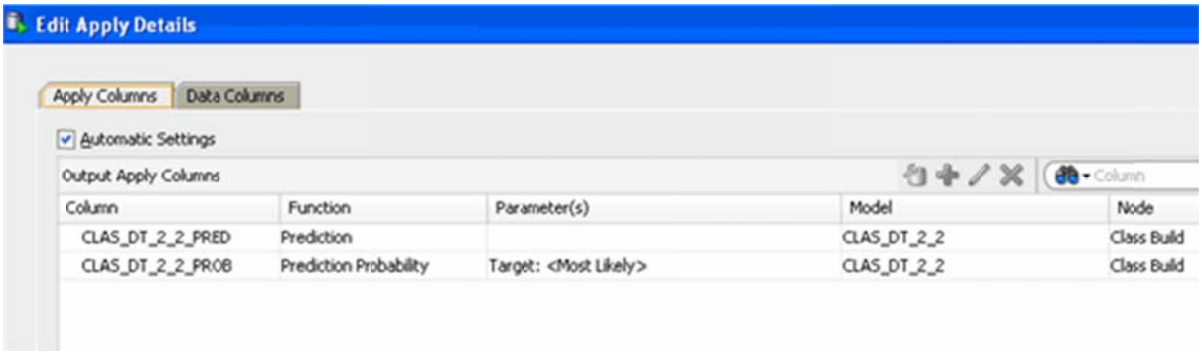
The prediction (Yes or No)
The probability of the prediction

However, you really want to know this information for *each* customer, so that you can readily associate the preditive information with a given customer.

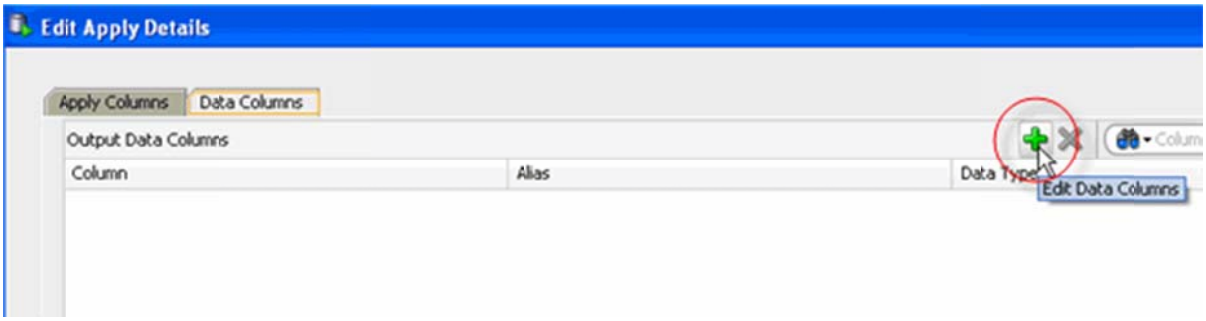
To get this information, you need to add a third column to the apply output: CUSTOMER_ID. Follow these instructions to add the customer id to the output:

A. Right-click the APPLY MODEL node and select **Edit**.

Result: The Edit Apply Details window appears. Notice that the Prediction and Probability columns are defined automatically in the Apply Columns tab.

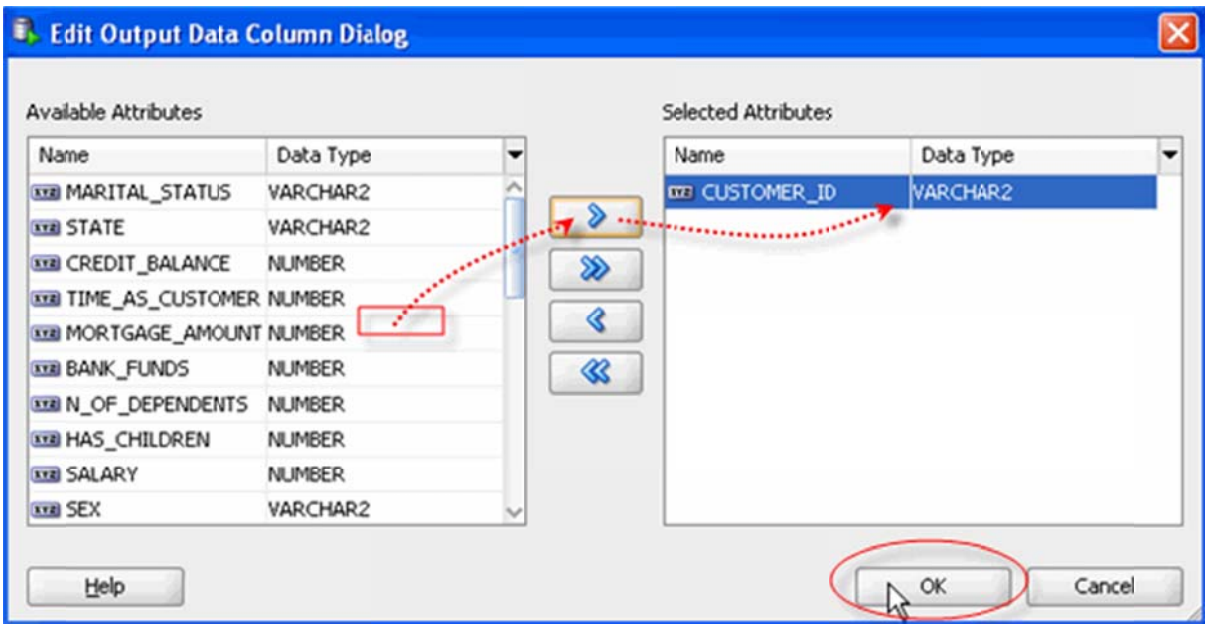


B. In the Edit Apply Details window, select the **Data Columns** tab, and then click the green "+" sign, like this:



C.In the Edit Output Data Column Dialog:

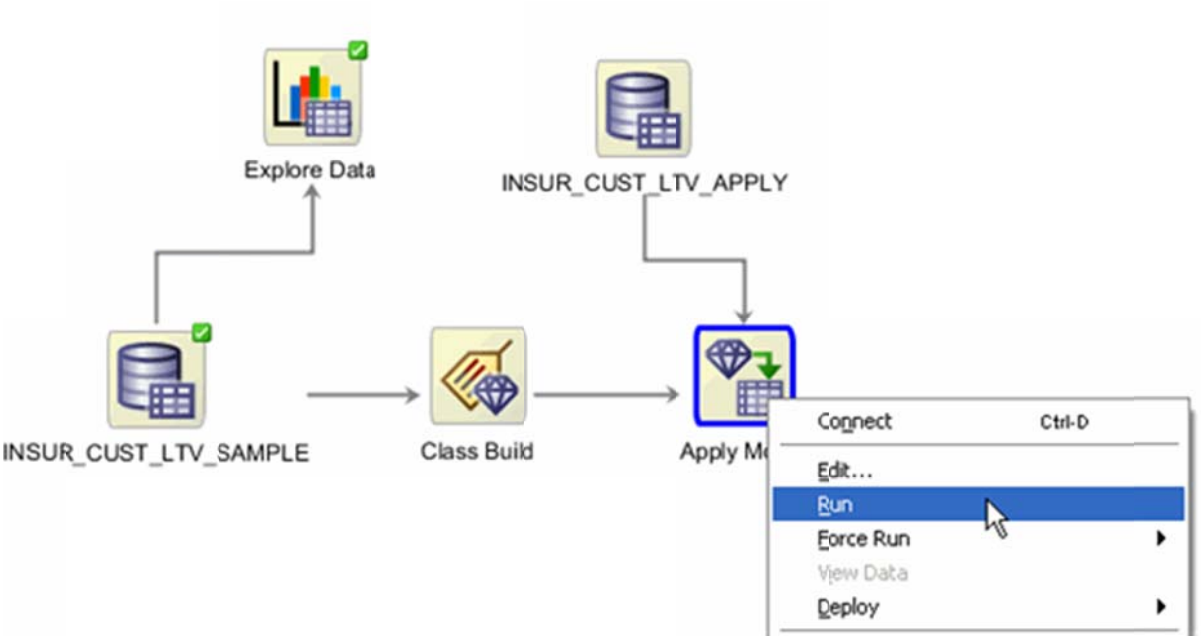
Select **CUSTOMER_ID** in the Available Attributes list
Move it to the Selected Attributes list by using the shuttle control
Then click **OK**.



Result: the CUSTOMER_ID column is added to the Data Columns tab.

D. Finally, click **OK** in the Edit Apply Details window to save the change.

8 . Now, you are ready to execute the Apply node. Right-click the Apply Model node and select **Run** from the menu.



Result:

As before, small green gear icons appear in each of the nodes that are being processed, and the Workflow Jobs tab displays the progress.
When the process is complete, green check mark icons are displayed in the border of all workflow nodes, and the

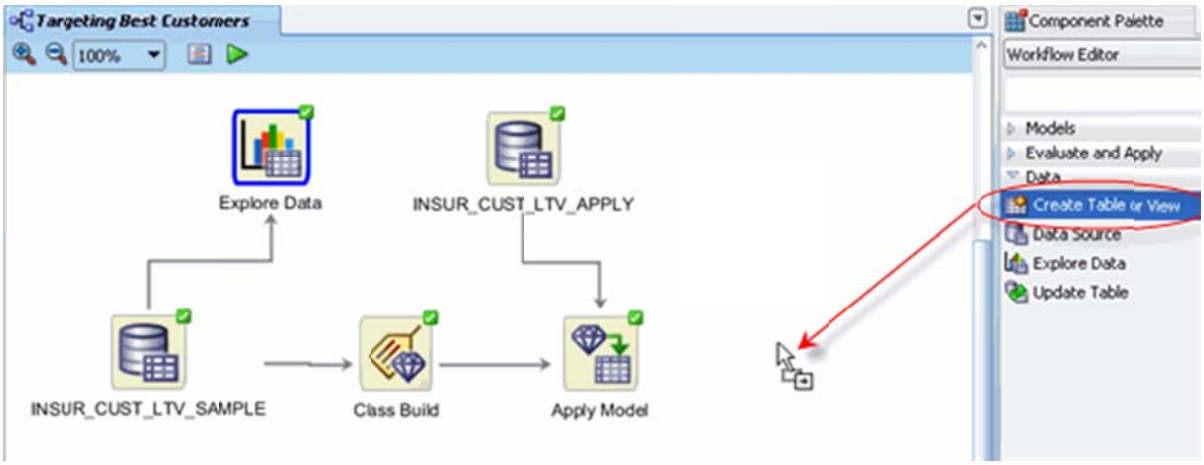
workflow document is automatically saved.

9 . Optionally, you can create a database table to store the model prediction results (the "Apply" results).

The table may be used for any number of reasons. For example, an application could read the predictions from that table, and suggest an appropriate response, like sending the customer a letter, offering the customer a discount, or some other appropriate action.

To create a table of model prediction results, perform the following:

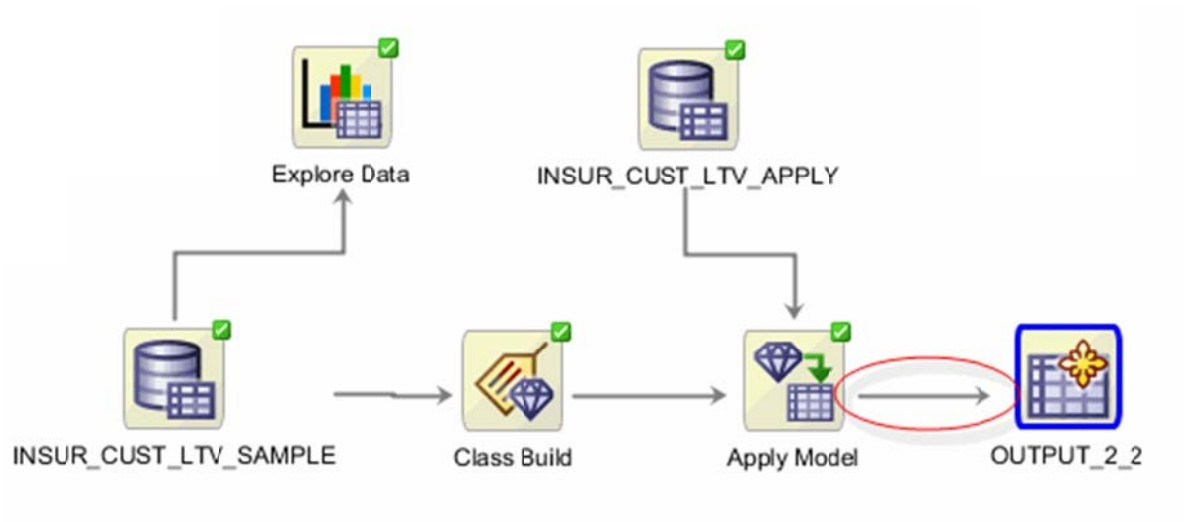
A. Using the Data list in the Component Palette, drag the **Create Table or View** node to the workflow canvas, like this:



Result: an OUTPUT node is created.

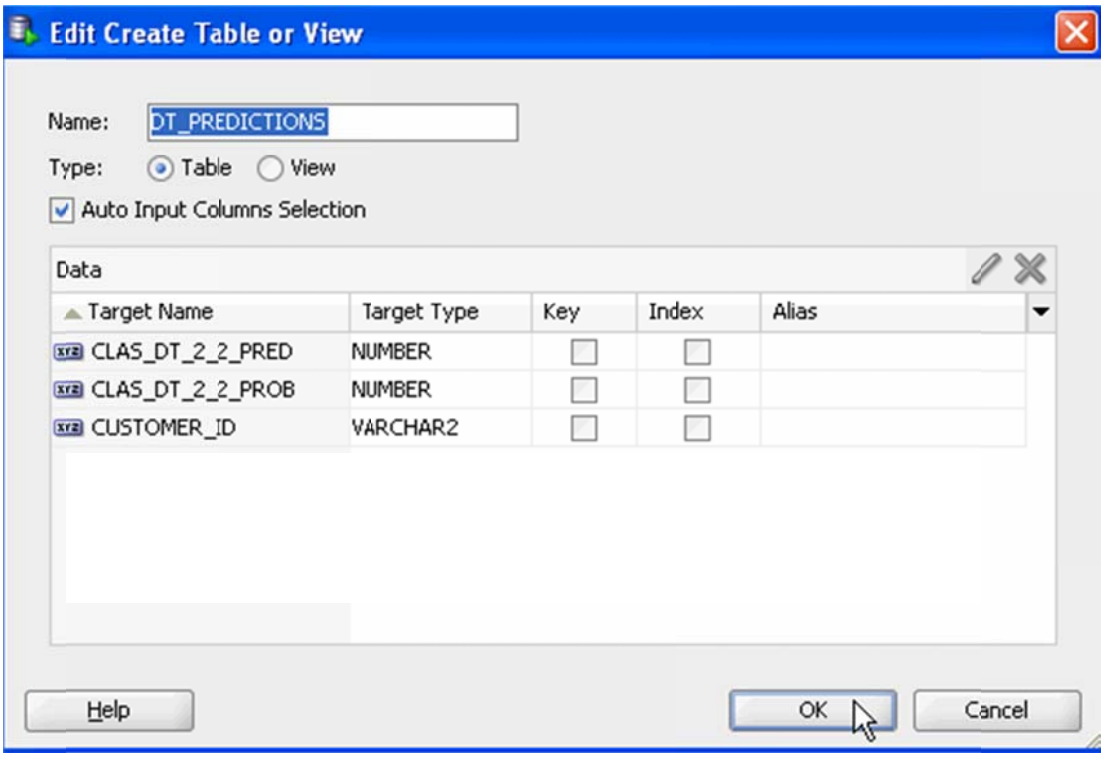
Required: Capture this screen and paste it onto your submission

B. Connect the **Apply Model** node to the **OUTPUT** node.



C. Specify a name for the table that will be created (otherwise, Data Miner will create a default name):

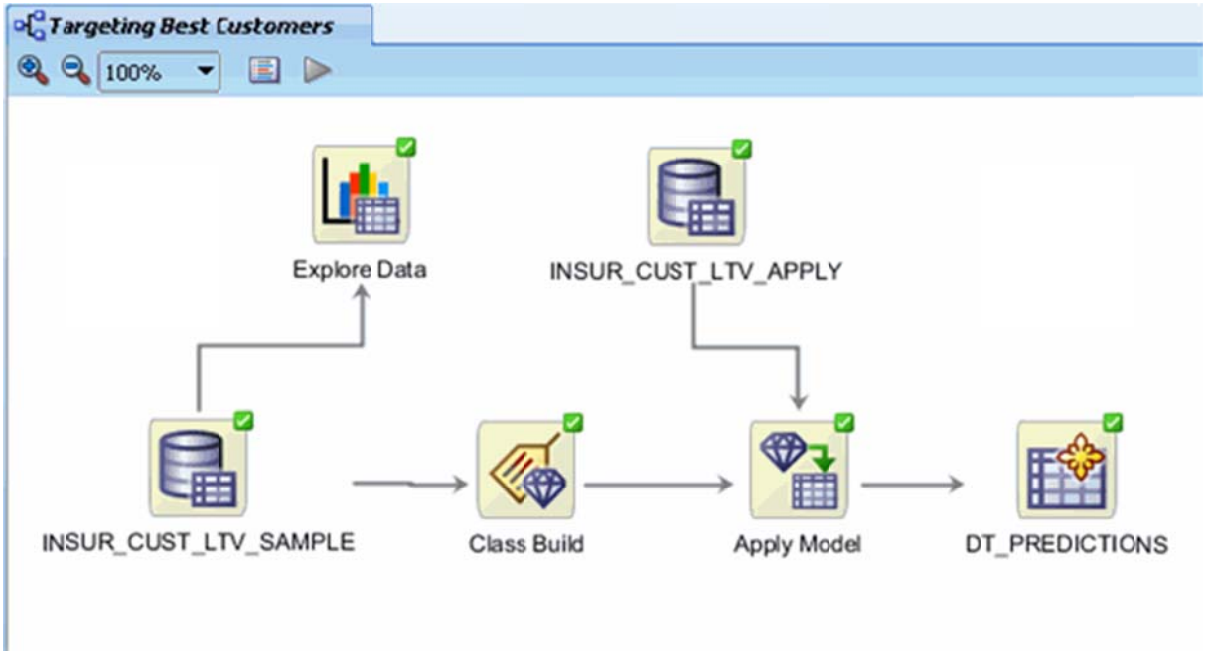
1. Right-click the OUTPUT node and select **Edit** from the menu.
2. In the Edit Create Table or View window, change the default table name to **DT_PREDICTIONS**, as shown here



3. Then, click **OK**.

D. Lastly, right-click the DT_PREDICTIONS node and select **Run** from the menu.

Result: When the process is complete, the workflow document is automatically saved, and all nodes contain a green check mark in the border, like this:



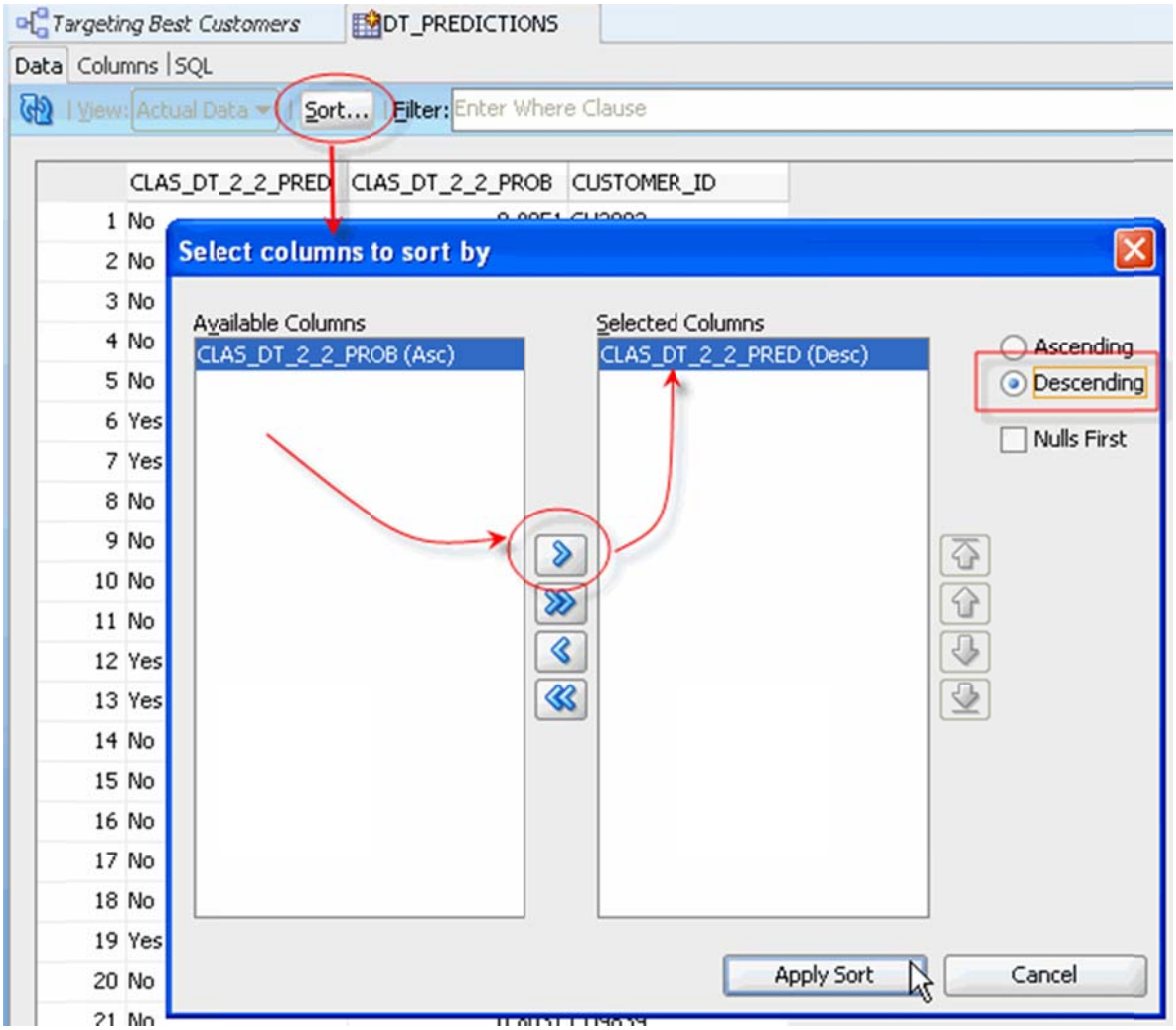
Note: After you run the OUTPUT node (DT_PREDICTIONS), the table is created in your schema.

10 . A. To view the results, right-click the DT_PREDICTIONS Table node and select **View Data** from the Menu.

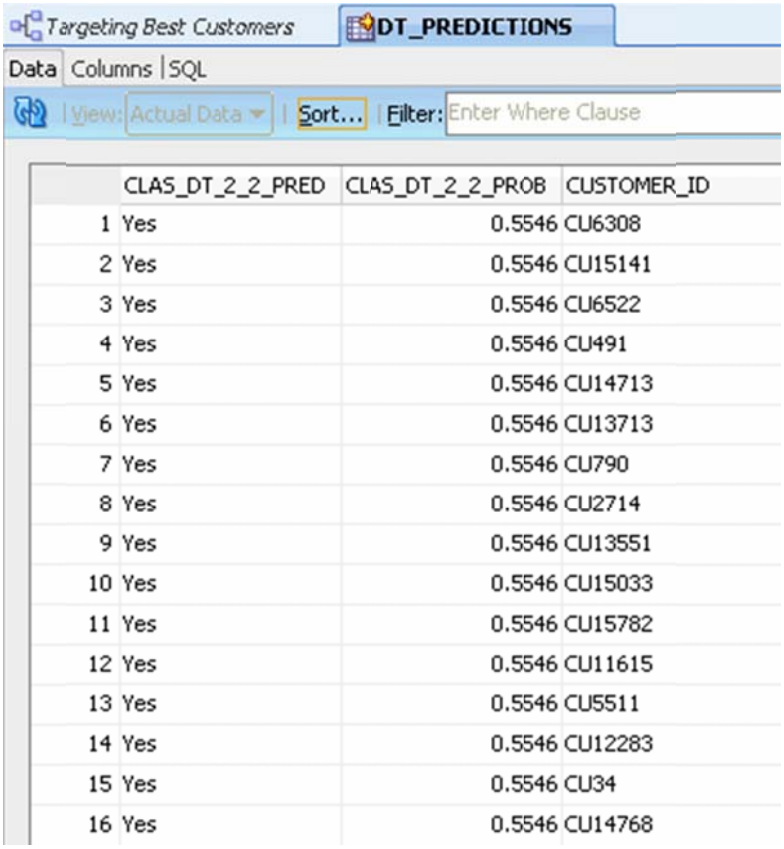
Result: A new tab opens with the contents of the table.

Notes:

The table contains two columns: one for the predicted outcome (Yes or No) and one for the probability of the prediction.
You can sort the table results on any of the columns using the Sort button, as shown here.
In this case, the table will be sorted using the Predicted outcome (**CLAS_DT_2_2_PRED** column), in**Descending** order.



B. Click **Apply Sort** to view the results:



	CLAS_DT_2_2_PRED	CLAS_DT_2_2_PROB	CUSTOMER_ID
1	Yes	0.5546	CU6308
2	Yes	0.5546	CU15141
3	Yes	0.5546	CU6522
4	Yes	0.5546	CU491
5	Yes	0.5546	CU14713
6	Yes	0.5546	CU13713
7	Yes	0.5546	CU790
8	Yes	0.5546	CU2714
9	Yes	0.5546	CU13551
10	Yes	0.5546	CU15033
11	Yes	0.5546	CU15782
12	Yes	0.5546	CU11615
13	Yes	0.5546	CU5511
14	Yes	0.5546	CU12283
15	Yes	0.5546	CU34
16	Yes	0.5546	CU14768

Required: Capture this screen and paste it onto your submission

Notes:

Each time you run an Apply node, Oracle Data Miner takes a different sample of the data to display. With each Apply, both the data and the order in which it is displayed may change. Therefore, the sample in your table may be different from the sample shown here. This is particularly evident when only a small pool of data is available, which is the case in the schema for this lesson.

You can also filter the table by entering a Where clause in the Filter box.

The table contents can be displayed using any Oracle application or tools, such as Oracle Application Express, Oracle BI Answers, Oracle BI Dashboards, and so on.

C. When you are done viewing the results, dismiss the tab for the DT_PREDICTIONS table, and click **Save All**.

Summary

In this lesson, you examined and solved a "Classification" prediction data mining business problem by using the Oracle Data Miner graphical user interface, which is included as an extension to SQL Developer, version 3.0.

In this tutorial, you have learned how to:

- Identify Data Miner interface components
- Create a Data Miner project
- Build a Workflow document that uses Classification models to predict customer behavior.