Table of Contents

Practices for Lesson 1: Introduction................................................................. 1-1
  Practices for Lesson 1: Overview................................................................. 1-2

Practices for Lesson 2: Data Mining Concepts and Terminology.................. 2-1
  Practices for Lesson 2: Overview................................................................. 2-2

Practices for Lesson 3: The Data Mining Process ........................................ 3-1
  Practices for Lesson 3: Overview................................................................. 3-2

Practices for Lesson 4: Introducing Oracle Data Miner 11g Release 2 .......... 4-1
  Practices for Lesson 4.................................................................................. 4-2
  Practice 4-1: Creating the Data Miner User Connection and Installing the Data Miner Repository ...... 4-3
  Practice 4-2: Creating a Data Miner Project and Workflow ......................... 4-7

Practices for Lesson 5: Using Classification Models...................................... 5-1
  Practices for Lesson 5.................................................................................. 5-2
  Practice 5-1: Selecting and Examining Source Data.................................... 5-3
  Practice 5-2: Creating Classification Models................................................ 5-7
  Practice 5-3: Scoring the Data..................................................................... 5-12

Practices for Lesson 6: Using Regression Models......................................... 6-1
  Practices for Lesson 6.................................................................................. 6-2
  Practice 6-1: Creating a New Workflow and Examining the Source Data ....... 6-3
  Practice 6-2: Creating Regression Models................................................... 6-8
  Practice 6-3: Scoring the Data..................................................................... 6-12

Practices for Lesson 7: Using Clustering Models............................................ 7-1
  Practices for Lesson 7.................................................................................. 7-2
  Practice 7-1: Creating Clustering Models................................................... 7-3
  Practice 7-2: Viewing Model Results and Applying the Model................. 7-6

Practices for Lesson 8: Performing Market Basket Analysis........................ 8-1
  Practices for Lesson 8.................................................................................. 8-2
  Practice 8-1: Creating a New Workflow and Examining the Source Data ....... 8-3
  Practice 8-2: Creating an Association Rules Model...................................... 8-7
  Practice 8-3: Modifying the Source Data and Rebuilding the Model............ 8-11
  Practice 8-4: Applying Model Rule Filters and Creating Model Details ...... 8-16

Practices for Lesson 9: Performing Anomaly Detection.................................. 9-1
  Practices for Lesson 9.................................................................................. 9-2
  Practice 9-1: Creating an Anomaly Detection Model.................................... 9-3
  Practice 9-2: Viewing Model Results and Applying the Model................. 9-6

Practices for Lesson 10: Deploying Data Mining Results............................... 10-1
  Practices for Lesson 10............................................................................. 10-2
  Practice 10-1: Creating a Table for Data Mining Results............................ 10-3
  Practice 10-2: Deploying the SQL Code for Data Mining Results .............. 10-9
  Practice 10-3: Modifying Classification Models and Redeploying the Code... 10-14
Preface

How This Course Is Organized
This is an instructor-led course featuring lectures and hands-on exercises. Online demonstrations and written practice sessions reinforce the concepts and skills introduced.
Related Publications

Oracle Publications

<table>
<thead>
<tr>
<th>Title</th>
<th>Part Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle Data Mining Concepts 11g Release 2 (11.2)</td>
<td>E16808-06</td>
</tr>
<tr>
<td>Oracle Data Mining Application Developer's Guide 11g Release 2 (11.2)</td>
<td>E12218-07</td>
</tr>
</tbody>
</table>

Additional Publications
- System release bulletins
- Installation and user’s guides
- Read-me files
- International Oracle User’s Group (IOUG) articles
- Oracle Magazine
Practices for Lesson 1: Introduction

Chapter 1
Practices for Lesson 1: Overview

Practices Overview
There are no practices for this lesson.
Chapter 2 - Page 1
Practices for Lesson 2: Overview

Practices Overview
There are no practices for this lesson.
Practices for Lesson 3: The Data Mining Process

Chapter 3
Practices for Lesson 3: Overview

Practices Overview
There are no practices for this lesson.
Practices for Lesson 4: Introducing Oracle Data Miner 11g Release 2

Chapter 4
Practices for Lesson 4

Practices Overview
In these practices, you create a connection for the Data Miner user, install the Data Miner repository, and create your first Data Miner objects: a project and a workflow.
Practice 4-1: Creating the Data Miner User Connection and Installing the Data Miner Repository

Overview
In this practice, you create a SQL Developer connection for the Data Miner user, display the user on the Data Miner tab, and install the Data Miner repository.

Tasks
First, create a SQL Developer connection for the Data Miner user.

1. Using the SQL Developer Connections tab, right-click the Connections node and select New Connection.

2. In the New / Select Database Connection dialog box, specify the following:
   - Connection Name: dmuser
   - Username: dmuser
   - Password: dmuser
   - Connection Type: Basic
   - Role: default
   - Host Name: localhost
   - Port: 1521
3. Still in the New / Select Database Connection dialog box, click Test to confirm that the connection is correct.
   **Result:** A status of “Success” should appear in the lower-left corner of the dialog box.
4. Click Save and then dismiss the dialog box.

Next, open the Data Miner tab and add a connection for the Data Miner user.

5. Open the Data Miner tab by selecting **Tools > Data Miner > Make Visible** from SQL Developers’ menu bar.

6. On the Data Miner tab, click the Add Connection icon ( ).
   **Result:** The Select Connection dialog box appears.
7. Select the Data Miner user (**dmuser**) from a list of existing connections and click OK.
Result: The dmuser connection appears on the Data Miner tab.

Finally, install the Data Miner repository.

8. To launch the installation routine, double-click dmuser on the Data Miner tab.

Result: A message tells you that the Data Miner repository is not installed in the database, and asks you whether you want to install the repository.

9. Click Yes to launch the installation process.

10. In the Connection Information dialog box, the SYS username is auto-filled. Enter oracle for the password, and click OK.

11. In the Repository Installation Settings dialog box, select the following values for the ODMRSYS account, and click OK.
Default Tablespace: **USERS**
Temporary Tablespace: **TEMP**

12. In the Install Data Miner Repository dialog box, click **Start** to begin the repository installation. Ensure that the Install Demo Data option is selected.

**Notes**
- The Installation will take several minutes to complete.
- The scroll bar provides a visual indication of the installation tasks and progress.

13. When the installation process is finished, click **Close** to dismiss the Install Data Miner Repository dialog box.
Practice 4-2: Creating a Data Miner Project and Workflow

Overview
In this practice, you create a Data Miner project and a workflow.

Tasks
1. Dismiss the SQL Developer Connections tab by clicking the Close icon (X) on the Connections tab.
2. On the Data Miner tab, right-click dmuser and select New Project.
3. In the Create Project dialog box, enter ABC Insurance as the name, optionally enter a comment for the project, and click OK.
4. On the Data Miner tab, right-click the ABC Insurance project and select New Workflow.
5. In the Create Workflow dialog box, enter Buy Insurance as the name, and click OK.

Result: The default layout of the Data Miner workflow interface opens.


Result: The Property Inspector pane automatically fills the space that the Messages – Log pane previously used.

In the next practice, you work on this data mining workflow.
Practices for Lesson 5

Practices Overview
In these practices, you build a workflow that uses classification models to predict which customers are most likely to purchase insurance.
Practice 5-1: Selecting and Examining Source Data

Overview
In this practice, you select and examine the source data for the Buy Insurance workflow that you created previously.

Assumptions
The Buy Insurance workflow is open in the Data Miner interface. If it is not open, simply double-click the workflow on the Data Miner tab.

Tasks
First, add a Data Source node to the workflow.

1. In the Workflow Editor of the Component Palette, select the Data category to view its four options.

2. Drag and drop a Data Source node from the Component Palette into the workflow window.

Result: Step 1 of the Define Data Source Wizard automatically opens.

3. In the Available Tables/Views list, select the INSUR_CUST_LTV_SAMPLE table, and then click Finish.

Result: A Data Source node with the table name appears in the workflow window.
Next, add an Explore Data node to the workflow.

4. Drag and drop an **Explore Data** node from the Component Palette onto the workflow.

![Explore Data Diagram](image)

5. Connect the Data Source node to the Explore Data node by doing the following:
   a. Right-click the **Data Source** node (INSUR_CUST_LTV_SAMPLE) and select **Connect**.
   b. Drag the pointer to the Explore Data node.
   c. Click the **Explore Data** node to connect the two nodes.

**Result:** The workflow should now look similar to the following screenshot:

![Workflow Diagram](image)

6. Edit the Explore Data node to specify a Group By selection by doing the following:
   a. Right-click the **Explore Data** node and select **Edit** to open the Edit Explore Data Node dialog box.
   b. In the Group By list, select the **BUY_INSURANCE** attribute.
c. Click **OK**.

7. Run the Explore data node by selecting **Run** from the node’s right-click menu.

**Result:** When the run process is complete, the workflow should look similar to the following screenshot:

8. Select **View Data** from the Explore Data node’s right-click menu to examine the source data.

9. Use the Statistics tab of the Explore Data window to browse source data attributes. Each attribute provides a histogram, and other statistical information, because they are related to the Group By attribute of **BUY_INSURANCE**.
10. When you are finished browsing the source data, dismiss the Explore Data window by clicking the Close (X) icon: Explore Data X.
Practice 5-2: Creating Classification Models

Overview
In this practice, you perform the following tasks:
- Define four classification models
- Train the models against the source data.
- Examine test results.

Assumptions
Practice 5-1 has been completed successfully.

Tasks
First, add a Classification node to the workflow and define the classification models.
1. Open the Models category in the Component Palette to view the available list of options.
2. Drag and drop the Classification node from the Component Palette to the workflow.

3. Connect the Data Source node to the Class Build node by doing the following:
   a. Right-click the Data Source node (INSUR_CUST_LTV_SAMPLE) and select Connect.
   b. Drag the pointer to the Class Build node.
   c. Click the Class Build node to connect the two nodes.

By default, when you click the Class Build node to complete the connection, Data Miner automatically opens the Edit Classification Build Node dialog box.

4. In the Edit Classification Build Node dialog box, select BUY_INSURANCE as the Target attribute, and CUSTOMER_ID as the Case ID attribute.
5. Click the **Advanced** button.

6. In the Advanced Settings dialog box, select the following:
   a. The **Support Vector Machine** algorithm in the Model Settings list
   b. The **Algorithm Settings** tab
   c. **Linear** from the Kernel Function list

7. Click **OK** to close the Advanced Settings dialog box.
8. Click OK to close the Edit Classification Build Node dialog box.

9. In the workflow window, ensure that the Class Build node is selected, and then, in the Property Inspector, do the following:
   a. Click the Test tab.
   b. Change the “Split for Test” value to 50.

The Property Inspector should look similar to the following screenshot:

Second, you train and test the models against the source data by running the Class Build node. Then, you can examine test results.

10. Right-click the Class Build node and select Run.

   **Result:** As the run process executes, Data Miner builds (trains) and tests all of the models that are defined in the node. The following status indicators show that execution is ongoing:
   - A green gear icon appears on all associated workflow node borders.
   - The Workflow Jobs tab shows the status of the build.

When the train and test processes are complete, the workflow window should look similar to the following screenshot:
In addition, the Property Inspector should show that all four models have successfully completed the build (train) and test phases. The Property Inspector should look something like the following screenshot:

![Class Build - Property Inspector](image)

<table>
<thead>
<tr>
<th>Models</th>
<th>Build</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Build</td>
<td>Test</td>
</tr>
<tr>
<td>CLAS_GLM_1_9</td>
<td>12/8/1...</td>
<td>12/8/1...</td>
</tr>
<tr>
<td>CLAS_SVM_1_9</td>
<td>12/8/1...</td>
<td>12/8/1...</td>
</tr>
<tr>
<td>CLAS_DT_1_9</td>
<td>12/8/1...</td>
<td>12/8/1...</td>
</tr>
<tr>
<td>CLAS_NB_1_9</td>
<td>12/8/1...</td>
<td>12/8/1...</td>
</tr>
</tbody>
</table>

**Note:** The model names will typically be different, because they are generated automatically.

11. Right-click the **Class Build** node and select **Compare Test Results**.

**Result:** The Class Build window opens, containing five tabs of comparative statistical information for the included models.

![Class Build window](image)

**Note:** Your actual test results may be somewhat different from the results shown here, because each train/test cycle against the same source data uses a different sample.
12. Using information from the Compare Test Results output (Class Build window), evaluate which of the four models provides the best test results, focusing specifically on:
   - Performance tab: Predictive Confidence % and Overall Accuracy %
   - Performance Matrix tab: Correct Predictions %
   - Lift tab: Lift Cumulative (between the 20th and 40th quantiles)

In our example, the Decision Tree model provides the best test results.

13. Dismiss the Class Build window.

14. In the workflow window, right-click the **Class Build** node and select **View Models > CLAS_DT_1_#** (the name for your Decision Tree model may be different).

   **Result:** The Decision Tree Model is displayed, using a graphical decision tree interface.

15. Using either the Thumbnail tab on the right, or the scrollbars in the primary tree window, navigate down to the last branch of the tree.

16. Select the node that contains the largest Confidence percent for a Yes prediction. In this example, Node 6 is selected. When you select a node in the tree, the If/Then rule appears on the Rule tab.

   ![Decision Tree Diagram](image)

   **Note:** The rule indicates a formula that the model will use to create a prediction for the target attribute. The formula includes the input attributes that are considered the most significant in predicting the target attribute (customers who are most likely to purchase insurance).

17. Dismiss the CLAS_DT_1_# window.
Practice 5-3: Scoring the Data

Overview
In this practice, you perform the following tasks:
- Select the best model (deselect unwanted models from the Class Build node).
- Add a new data source to the workflow for scoring purposes.
- Add an Apply node to the workflow and modify the scoring output format.
- Score the data and examine predictive results.

Assumptions
Practice 5-2 has been completed successfully.

Tasks
In our example, the Decision Tree (DT) model is chosen for scoring.
1. Select the Class Build node in the workflow window to view its properties.
2. In the Property Inspector, deselect each of the models except the CLAS_DT model:
   a. Click the Models tab in the Property Inspector.
   b. Click the green arrow (→) in the Output column of each of the following models: CLAS_GLM, CLAS_SVM, and CLAS_NB. This action changes the icon to a green arrow followed by a red “x” (❌). When you are finished, the Property Inspector should look similar to the following screenshot:

![Property Inspector Screenshot](image)
3. Drag and drop a second **Data Source** node to the workflow.

![Workflow Diagram](image1.png)

4. In step 1 of the Define Data Source Wizard, select the **INSUR_CUST_LTV_APPLY** table and click **Finish**.

5. Rename the new Data Source node to **INSUR_CUST_LTV_SCORE** either by using the Details tab of the Property Inspector, or by editing the node name directly in the workflow. **Result:** The workflow window should look similar to the following screenshot:

![Workflow Diagram](image2.png)

6. Drag an **Apply** node from the Component Palette “Evaluate and Apply” category to the workflow.

![Workflow Diagram](image3.png)

7. Create two connections to the Apply node.
   a. Connect **INSUR_CUST_LTV_SCORE** to Apply, by doing the following:
1) Right-click the **INSUR_CUST_LTV_SCORE** node and select **Connect**.
2) Drag the **pointer** to the Apply node.
3) Click the **Apply** node to connect the two nodes.

b. Connect Class Build to Apply, by doing the following:
1) Right-click the **Class Build** node and select **Connect**.
2) Drag the **pointer** to the Apply node.
3) Click the **Apply** node to connect the two nodes.

When you are finished with the connections, the workflow should look similar to the following screenshot:

The workflow is ready for scoring. However, before executing the scoring process, consider the resulting output.

In addition to the default predictive results, it would be helpful to add customer identification information to the output. This addition will enable you to associate the predictive information with a given customer.

8. To add the CUSTOMER_ID column to the output, perform the following:
   a. Double-click the **Apply** node to open the Edit Apply Node dialog box.

   **Result:** Three output columns are selected by default on the Apply Columns tab, including: Prediction (Yes or No), Prediction Probability, and Prediction Cost.

   b. Click the **Data Columns** tab, and then click the **Edit Data Columns** icon ( ).
c. In the Edit Output Data Column dialog box, move the CUSTOMER_ID attribute from the Available list to the Selected list.

![Edit Output Data Column Dialog Box]

- CUSTOMER_ID

- Available list
- Selected list

- Name: CUSTOMER_ID
- Data Type: VARCHAR2

---

d. Click OK in the Edit Apply Node dialog box to save the modified output format.

Finally, score the data and examine the predictive results.

9. In the workflow, right-click the Apply node and select Run (or, select the Apply node and click the Run icon at the top of the workflow window).

**Notes**
- As the Apply process executes, Data Miner scores the source data using the selected model.
- The following status indicators indicate that execution is ongoing:
  - A green gear icon appears on all associated workflow node borders.
  - The Workflow Jobs tab shows the status of the build.

When the scoring process is complete, the workflow window should look similar to the following screenshot:

![Workflow Screenshot]

10. To view the predictive results, right-click the Apply node and select View Data.

**Result:** An Apply Model window opens with the predictive results.
11. Sort the output by Prediction and then by Probability (both in descending order) by doing the following:
   
   a. Click **Sort**.
   b. In the “Select columns to sort by” dialog box, select the following sort columns:
      - CLAS_DT_#_#_PRED (Desc)
      - CLAS_DT_#_#_PROB (Desc)
   c. Click **Apply Sort** to execute the sort.

   **Result:** The output is sorted. The following screenshot shows some of the output:

12. When you are done viewing predictive results, dismiss the Apply Model window.

13. Save the workflow and close the workflow window.
Practices for Lesson 6: Using Regression Models

Chapter 6
Practices for Lesson 6

Practices Overview
In these practices, you create a workflow by using regression models to predict the lifetime value of existing customers.
Practice 6-1: Creating a New Workflow and Examining the Source Data

Overview
In this practice, you create a new workflow. Then you select and examine the source data for the regression modeling case study.

Assumptions
The ABC Insurance project has been created.

Tasks
1. Create a new workflow in the exiting ABC Insurance project, by doing the following:
   a. Right-click the existing project (ABC Insurance) and select New Workflow.
   b. In the Create Workflow dialog box, enter Predicting LTV as the name and click OK.

2. In the Workflow Editor of the Component Palette, select the Data category to view its four options.
3. Drag a Data Source node into the workflow window.
   Result: Step 1 of the Define Data Source Wizard automatically opens.
4. In the Define Data Source Wizard, select the CUST_INSUR_LTV table and click Finish.
   Result: A data source node with the table name appears in the workflow window.
Next you add an Explore Data node to the workflow.

5. Drag an **Explore Data** node from the Component Palette into the workflow window.

6. Connect the data source node to the Explore Data node by doing the following:
   a. Right-click the data source node (**CUST_INSUR_LTV**) and select **Connect**.
   b. Drag the pointer to the Explore Data node.
   c. Click the **Explore Data** node to connect the two nodes.

7. Right-click the **Explore Data** node and select **Run**.

**Note:** In this case, you are not specifying a Group By attribute for the Explore Data node.

8. Once the run process completes, edit the Explore Data node to view statistics for the source data. To begin editing the node, do either of the following:
   - Double-click the node.
   - Right-click the node and select **View Data**.

9. On the Statistics tab of the Explore Data window, scroll down the list of attributes and select the **LTV** attribute.
Notes

- The LTV attribute will be used as the Target attribute for this modeling exercise. It will contain the predicted lifetime value amount for each customer.

- All other attributes that are included in the source data are used as input variables. In other words, the regression model uses them as input in order to predict the LTV for each customer.

- However, the LTV_BIN attribute (shown just below the LTV attribute), is not considered an input variable for a regression model. Instead, it is used for Classification modeling purposes, because it stores values such as HIGH, MEDIUM, and LOW, which categorize customers into a certain level of value. Therefore, this attribute must be filtered out of the source data.

10. To filter the LTV_BIN column out of the source data, do the following:
   a. Open the **Transforms** node category in the Component Palette.
   b. Drag the **Filter Columns** node into the workflow window.
c. Connect the data source node to the Filter Columns node.
d. After you connect the nodes, right-click the Filter Columns node and select Edit to specify the filter criteria.

e. In the Edit Filter Columns Node dialog box, deselect the LTV_BIN column by clicking the green arrow (➡️) in the Output column for that attribute. The result should look similar to the following screenshot:
f. Click **OK** to save the filter specification.

11. Run the Filter Columns node by doing either of the following:
   - Select **Run** from node’s right-click menu.
   - Select the node and click the **Run icon** (▶) at the top of the workflow window.

**Note:** The run process records the filtering instruction in the workflow document, preventing the LTV_BIN column from being passed forward as part of the source data.

When the run process is complete, the workflow should look similar to the following screenshot:
Practice 6-2: Creating Regression Models

Overview
In this practice, you perform the following tasks:
- Define two regression models.
- Build the models against the source data and examine test results.

Assumptions
Practice 6-1 has been completed successfully.

Tasks
You first add a Regression node to the workflow and define the regression models by doing the following:
1. Open the Models category in the Component Palette to view the available list of options.
2. Drag the Regression node from the palette into the workflow window.

   Result: The workflow window should look similar to the following screenshot.

   ![Workflow Diagram](image)

   **Note:** Regress Build node validation messages appear when you place the cursor over the information warning (yellow “!”) icon.

3. Connect the Filter Columns node to the Regress Build node.

   By default, when you click the Regress Build node to complete the connection, Data Miner automatically opens the Edit Regression Build Node dialog box.

4. In the Edit Regression Build Node dialog box do the following:
   a. Select LTV as the Target attribute.
   b. Select CUST_ID as the Case ID attribute.
c. Click OK to save the build node definition.

Now, train and test the models against the source data by running the Regress Build node. Then you can examine test results.

5. Right-click the **Regress Build** node and select **Run**.

**Results:**
- As the run process executes, Data Miner builds (trains) and tests the two models that are defined in the node.
- When the process is complete, all nodes in the workflow window contain green check-marks in the node border, and the Property Inspector shows that both models have successfully completed the build (train) and test phases.
Note: The model names will typically be different, because they are generated automatically.

6. Right-click the **Regress Build** node and select **Compare Test Results**.

**Result:** The Regress Build window opens, containing two tabs of statistical information for the model: Performance and Residual.

7. Click the **Residual** tab. The default presentation of test results should look similar to the following screenshot:

---

**Notes**

- By default:
  - The first model is displayed in the graph (selected in the Show drop-down list). In this case, the GLM model is displayed.
  - Residual values are displayed on the Y-axis and predicted values are displayed on the X-axis.
- Your actual test results may look slightly different, because each train/test cycle against the same source data uses a different sample.
8. On the Residual tab, change the following options:
   - Y-Axis: **Predicted**
   - X-Axis: **Actual**
   - Compare to: **REGR_SVM_#_#**

   **Note:** The REGR_GLM model is already selected in the Show drop-down list.

9. Click the **Query** button to refresh the results.

   ![Diagram showing residual plots for REGR_GLM_1_10 and REGR_SVM_1_10.]

   **Notes**
   - The comparative view provides evidence of the difference in the fit of each model. In the screenshot above, the GLM model provides a tighter fit than the SVM model when comparing the predicted and actual values.
   - Your results may look slightly different, but the overall result should be similar.
   - Therefore, the GLM model will be selected for scoring.

10. Dismiss the Regress Build window by clicking the **Close (X)** icon on the tab.
Practice 6-3: Scoring the Data

Overview
In this practice, you perform the following tasks:
- Select the best model.
- Add a new data source to the workflow for scoring purposes.
- Add an Apply node to the workflow and modify scoring output format.
- Score the data and examine predictive results.

Assumptions
Practice 6-2 has been completed successfully.

Tasks
In our example, the GLM model is chosen for scoring.
1. Select the Regress Build node in the workflow window to view its properties.
2. In the Property Inspector, deselect the SVM model by doing the following
   a. Click the Models tab in the Property Inspector.
   b. Click the green arrow ( ) in the Output column of the REGR_SVM model
      The Property Inspector should now look similar to the following screenshot:

3. Add a second Data Source node to the workflow, just above and to the right of the Regress Build node.
Result: The Define Data Source Wizard automatically opens.

4. In Step 1 of the Define Data Source Wizard, select the **CUST_INSUR_LTV APPLY** table and click **Next**.

5. In Step 2 of the Define Data Source Wizard, move both the **LTV** and **LTV BIN** attributes from the Selected Attribute list to the Available Attributes list.

6. Click **Finish** to save the new data source node definition.

   **Note:** In this way, both of these target attributes are removed from the scoring process, and only the input variables remain.

7. Rename the new data source node to **CUST_INSUR_LTV_NEW** either by using the Details tab of the Property Inspector, or by editing the node name directly in the workflow.

**Result:** The workflow window should look similar to the following screenshot:
8. Drag and drop an **Apply** node from the Component Palette “Evaluate and Apply” category to the workflow.

9. Create two connections to the Apply node, as follows:
   a. First:
      1) Right-click the **CUST_INSUR_LTV_NEW** node and select **Connect**.
      2) Drag the **pointer** to the Apply node.
      3) Click the **Apply** node to connect the two nodes.
   b. Second:
      1) Right-click the **Regress Build** node and select **Connect**.
      2) Drag the **pointer** to the Apply node.
      3) Click the **Apply** node to connect the two nodes.

When you are finished with the connections, the workflow should look similar to the following screenshot:
The workflow is ready for scoring. However, as you saw with the classification model workflow, it is helpful to consider the resulting output before executing the scoring process.

By default, a regression model includes both Lower- and Upper-Bounds Prediction Confidence percentage parameters in addition to the Prediction column. In this example, you remove the two bounds columns, and add several data columns to the output.

10. To modify the default output, do the following:
   a. Double-click the **Apply** node to open the Edit Apply Node dialog box.
      
      **Result:** The three default prediction columns are shown on the Apply Columns tab of the Edit Apply Node window.
   b. Deselect the **Automatic Settings** option at the top of the tab.
   c. Select the two **Prediction Bounds** columns and click the **Remove Apply Columns icon** (red “X” icon).

   **Result:** Only the Prediction column remains on the Apply Columns tab.
   d. Click the **Data Columns** tab, and then click the **Edit Data Columns icon**.
e. In the Edit Output Data Column dialog box, move the **CUST_ID** attribute from the Available list to the Selected list.

f. Add the following input variables to the Selected list:
   - **HOUSE_OWNERSHIP**
   - **MORTGAGE_AMOUNT**
   - **TIME_AS_CUSTOMER**
   - **SALARY**

**Result:** The Edit Output Data Column Dialog should look like similar to the following screenshot:

![Edit Output Data Column Dialog](image)

**Note:** Other input variables that are believed to be significant contributors to the target attribute (lifetime value) may also be added to the output.

g. Click **OK** to save the selected output data columns.
h. Change the Default Column Order option to **Apply Columns First**.

i. Click **OK** in the Edit Apply Node window to save the modified output format.
To complete this practice, score the data and examine the predictive results.

11. In the workflow, right-click the **Apply** node and select **Run**.

When the scoring process is complete, the workflow window should look similar to the following screenshot:

12. To view the predictive results, right-click the **Apply** node and select **View Data**.

**Result:** An Apply window opens with the predictive results and those data columns included in the output.

13. In the Apply window, click the **Sort** tool and specify the following sort criteria: First sort by the prediction (descending order), and then by customer ID (ascending order).

14. Click **Apply Sort** to execute the sort and view the resulting output.
15. If so desired, you can modify the Apply node to include other input attributes in your output, and then re-execute the scoring process by selecting Run from the Apply node’s right-click menu.

16. When you are done viewing predictive results, dismiss the Apply window.

17. Save and close the workflow.

**Note:** Your sort results may differ slightly.
Practices for Lesson 7: Using Clustering Models

Chapter 7
Practices for Lesson 7

Practices Overview
In these practices, you create a workflow that uses clustering models to discover the different segments, or groups, into which existing customers naturally fit.
Practice 7-1: Creating Clustering Models

Overview
In this practice, you create a new workflow, specify the source data, and define clustering models for the case study. Because this new workflow will use the same source data as the Predicting LTV workflow, you reuse the data source node from that workflow.

Assumptions
Practice 6-1 has been successfully completed.

Tasks
1. To create a new workflow in the exiting ABC Insurance project, do the following:
   a. Right-click the ABC Insurance project and select New Workflow.
   b. In the Create Workflow dialog box, use Customer Segmentation as the name and click OK.
2. Open the Predicting LTV workflow from the Data Miner tab. Both workflow tabbed panes are now available.
3. In the Predicting LTV workflow window, select the CUST_INSUR_LTV node and copy it to the clipboard.
4. Switch to the Customer Segmentation workflow window and paste the clipboard contents.

Result: The CUST_INSUR_LTV node appears in the Customer Segmentation workflow.

5. Close the Predicting LTV workflow by clicking the Close icon (X) on its tab.

6. Open the Models category in the Component Palette and drag the Clustering node from the Component Palette into the Customer Segmentation workflow window.
7. Rename the Clust Build node Cluster Build by using either the Property Inspector or by editing the node name directly in the workflow.

8. Connect the data source node to the Cluster Build node.

9. Right-click the Cluster Build node and select Edit.

10. In the Edit Clustering Build Node dialog box, select CUST_ID as the Case ID attribute and then click OK.

   ![Edit Clustering Build Node dialog box](image)

   **Note:** The default settings for both algorithms are used.
11. Run the Cluster Build node by either selecting **Run** from the node’s right-click menu, or by selecting the node and clicking the Run icon (▶).

When the build process is complete, the workflow window should look similar to the following screenshot:
Practice 7-2: Viewing Model Results and Applying the Model

Overview
In this practice, you view the results for the K-Means model and apply that model to the data.

Assumptions
Practice 7-1 has been successfully completed.

Tasks
1. Right-click the Cluster Build node and select View Models > CLUS_KM_#. In the resulting tree structure display, click Cluster1 at the top of the tree, and then select the Centroid tab.

Notes
- The model contains 10 leaf clusters, as specified by the K-Means algorithm setting for “Number of Clusters.”
- The Centroid tab displays a list of contributing attributes for the cluster, sorted by the Importance attribute.
- The Cluster Rule tab shows the rule for the selected cluster in If/Then format.
2. Navigate down to and select terminal node **Cluster 9**. Then click the **Centroid** tab.

**Notes**

Your model results may be slightly different. In our results:

- LTV and TIME_AS_CUSTOMER are the most significant attribute predictors of this cluster, with an Importance value of 0.6.
- Several other contributing attributes are ranked with the same Importance value of 0.5.
- Even though the LTV attribute is the most significant factor in the definition of this cluster, the lifetime value amounts of customers in this cluster are only in a medium range. We can see this in the Mean and Variance values for LTV, and also in the Mode value for the LTV_BIN attribute.
3. Navigate down in the tree and select terminal node **Cluster 11**. Then, click the **Centroid** tab.

![Diagram of tree with clusters and attributes]

**Cluster 10**
- Count: 287
- Percent: 1.87%
- Attributes: N_TRANS_ATM, N_TRANS_WEB_BANK, MORTGAGE_AMOUNT, HAS_CHILDREN, BUY_INSURANCE

**Cluster 11**
- Count: 1192
- Percent: 7.77%
- Attributes: N_OF_DEPENDENTS, MORTGAGE_AMOUNT, MARITAL_STATUS, CREDIT_CARD_LIMITS, LTV

**Cluster 13**
- Count: 250

**Notes**
- In our results, number of dependents is ranked highest as a cluster predictor, with an Importance value of 0.75. Members of this cluster have one or two children.
- Number of mortgages, LTV_BIN, house ownership, and bank funds are also significant attribute predictors of this cluster, with an Importance value of 0.6667. These attributes indicate a cluster of high-value customers. We can see that members of this cluster own one or more homes, with one or more mortgages.
- It is also notable that the mean value for LTV in Cluster 11 is significantly higher than in Cluster 9.
4. Click the **Detail** tab at the top of the window, and then select Cluster **11**.

5. In the resulting window, select the **LTV_BIN** attribute.

**Notes**

- The Details tab enables you to visualize each attribute of the cluster with a histogram.
- The contributing attributes for the cluster are sorted using the same mechanism as seen on the Tree – Centroid tab.
- The **LTV_BIN** attribute shows that the vast majority of customers in this cluster are predicted as either VERY HIGH or HIGH value customers.
6. Select the **LTV** attribute. The histogram shows the lifetime value contributions.

![Histogram showing lifetime value contributions](image)

7. Click the **Compare** tab. Then select **11** in the Cluster 1 box, and select **9** in the Cluster 2 box.

8. When the results refresh, click the **Edit** button next to Cluster 11.

![Rename Cluster dialog box](image)

9. In the Rename Cluster dialog box, enter **11 - High Value Customers**, and click **OK**.

10. Dismiss the CLUS_KM_#_# window.

Next you apply the K-Means model to the source data. You first deselect the O-Cluster model, and then you add and run an Apply node.

11. Select the **Cluster Build** node in the workflow.

12. Deselect the O-Cluster model by using the Property Inspector, as follows:
   a. Click the **Models** tab in the Property Inspector.
   b. Click the **green arrow** (➜) in the Output column of the CLUS_OC_#_# model, changing the icon to a green arrow followed by a red “x” (✗).
The Property Inspector should look similar to the following screenshot:

![Property Inspector Screenshot]

**Note:** The green check mark in the border of the Cluster Build node disappears, indicating that the node has been modified.

13. In the Component Palette, select the “Evaluate and Apply” category, and drag an **Apply** node to the workflow.

![Workflow Diagram]

14. Create the following two connections:
   a. Connect the data source (**CUST_INSUR_LTV**) node to the **Apply** node.
   b. Connect the **Cluster Build** node to the **Apply** node.

The workflow should now look like similar to the following screenshot:

![Workflow Diagram]

15. As you have learned, you can remove or add columns to the model output by editing the **Apply node**. Double-click the **Apply Node**, or select **Edit** from the node’s right-click menu.
16. In the Edit Apply Node dialog box, do the following:
   a. On the Apply Columns tab, select **Apply Columns First** as the Default Column Order.
   b. On the Data Columns tab, click the **Edit Data Columns icon** ()
   c. In the Edit Output Data Columns dialog box, move all data columns to the Selected Attributes list.
   d. Click **OK** to close the Edit Output Data Columns dialog box.
   e. Click **OK** in the Edit Apply Node dialog box to save the model output format.

17. Run the Apply node by either selecting **Run** from the node’s right-click menu, or by selecting the node and clicking the Run icon ().

18. When the apply process is complete, select **View Data** from the Apply node’s right-click menu.
19. As with all Apply data output, you can sort the results. Sort the output by the Probability column, in descending order.
20. Enter the following SQL WHERE clause in the Filter box:

```sql
CLUS_KM_1_35_CLID = '11 - High Value Customers'
```

**Notes**

- The name of your cluster ID (…_CLID) column will probably be different. Use the exact name of your cluster ID column.
- Some of the results of the WHERE clause filter are shown in the following screenshot. Certain columns have been resized.

Results: All of the customer records in this filtered output belong to the High Value Customers cluster (cluster number 11) that was discovered by the K-Means clustering model.

21. When you are done viewing the Apply results, close the Apply window, save the workflow, and close the workflow window.
Practices for Lesson 8: Performing Market Basket Analysis

Chapter 8
Practices for Lesson 8

Practices Overview
In these practices, you create a workflow that uses association rules models to perform market basket analysis of retail customers.

During the workflow creation process, you:

- Identify and examine sales data from a new schema
- Use the Association Rules algorithm to build a model from the sales data
- View and evaluate the model results for purposes of market basket analysis
Practice 8-1: Creating a New Workflow and Examining the Source Data

Overview
In this practice, you create a new project and a workflow, and then select a data source from a new schema. Then you examine the source data for the association rules modeling case study.

Tasks
1. On the Data Miner tab, right-click the dmuser connection and select New Project.
2. In the Create Project dialog box, enter Retail as the name and then click OK.
3. Create a new workflow in the Retail project, by doing the following:
   a. Right-click the Retail project and select New Workflow.
   b. In the Create Workflow dialog box, enter Market Basket Analysis as the name and click OK. The workflow window opens.
4. Drag a **Data Source** node from the Component Palette to the workflow window.

**Result:** Step 1 of the Define Data Source Wizard opens automatically.

5. In the wizard, click the **Add Schemas** button.

**Result:** the Edit Schema List dialog box opens.

6. In the Edit Schema List dialog box, move the **SH** schema from the Available Schemas list to the Selected Schemas list.

7. Click **OK**.

**Result:** The tables and views in the SH schema are now available to be used as data sources for mining activities.
8. Return to Step 1 of the Define Data Source Wizard, and do the following:
   a. Select the "Include Tables from Other Schemas" option.
      **Result:** Additional tables (and views) in the schema that you selected are displayed in the Available Tables/Views list. Tables and views are now prefixed with the schema name.
   b. Scroll down in the Available Tables/Views list and select the **SALES** table from the SH schema.

   ![Select Table Diagram]

   c. Click **Finish**.

9. To preview the source data, right-click the **SALES** node and select **View Data**.

   ![SALES Node Menu]

   **Connect** | **Ctrl-D**
   --- | ---
   Edit... | Attributes...
   Define Data Wizard... | Run
   **View Data** | Deploy
   Generate Apply Chain
10. In the resulting SH.SALES window, do the following:
   a. Click the **Sort** icon, and specify the following sort criteria:
      - CUST_ID (Ascending)
      - TIME_ID (Ascending)
   b. Click **Apply Sort**.

   **Result:** A sorted display of data provides a simplistic “basket” view of products that are purchased at a certain time by certain customers.

   **Note:** The order of columns in your output may differ, but the data will be the same.

11. Close the SH.SALES window by clicking the **Close icon (X)**.
Practice 8-2: Creating an Association Rules Model

Overview
In this practice, you define an association rules model to create a data mining prediction of which products will commonly sell together. You perform the following tasks:

- Define an association rules model.
- Build the model against the source data, and examine the model results.

Because association rules models are an unsupervised data mining technique, there is no separate scoring phase in the model building process.

Assumptions
Practice 8-1 has been completed successfully.

Tasks
You first add an Association node to the workflow to define the model settings.

1. Open the Models category in the Component Palette to view the available list of options.
2. Drag the Association node from the Component Palette to the workflow window.

Result: The workflow window should look similar to the following screenshot.

Note: Assoc Build node validation messages appear when you place the cursor over the information warning (yellow “!”) icon.

3. Connect the data source node to the Assoc Build node.
By default, when you click the Assoc Build node to complete the connection, Data Miner automatically opens the Edit Association Build Node dialog box.

The Edit Association Build Node dialog box should now look similar to the following screenshot:

In this dialog box, at least two options must be defined, as indicated by the validation indicators.

4. To define a value for the Transaction IDs option, do the following:
   a. Click the Select button to the right of the Transaction IDs box.
   b. In the Select Columns dialog box, move the CUST_ID and TIME_ID attributes from the Available Attributes list to the Selected Attributes list.
   c. Click OK.

5. For the Item ID option, select PROD_ID.

The Edit Association Build Node dialog box should now look similar to the following screenshot:
Notes

- The **Item ID** option specifies the attribute that you want bundled (Product).
- The **Transaction IDs** option specifies how you want to define the “basket.” The combination of Customer and Time determines those products that are sold together for each customer at one shopping visit.
- The **Value** option lets you to specify another column within the transaction data to combine with the Item ID. A value of <Existence> means that you want to see whether there is any type of common bundling among all values for the selected Item ID.

6. Click the **Advanced** button to open the Advanced Settings dialog box.
7. Change the Maximum rule length from its default value to **3**.
8. Click OK to save the modified advanced setting.

9. Click OK in the Edit Association Build Node dialog box to save the model definition.

10. Right-click the Assoc Build node and select Run, to build the model.

11. When the build process is complete, select View Models > ASSOC_AP from the Assoc Build node’s right-click menu.

Result: An ASSOC_AP window opens with the Rules tab displayed by default.
Note: The rule ID values are generated automatically by Data Miner. Your ID values may be different. As seen in the screenshot above, the rule content and rule details panes use product ID values for Antecedent and Consequent values.

12. Close the ASSOC_AP window.
Practice 8-3: Modifying the Source Data and Rebuilding the Model

Overview
In this practice, you modify the source data to include product descriptions. Then you rebuild the model and examine the results.

Assumptions
Practice 8-2 has been completed successfully.

Tasks
In order to view the model rules with product descriptions, rather than ID codes, join the SALES and PRODUCTS tables.

1. Add a new Data Source node beneath the SALES node. In Step 1 of the Define Data Source Wizard, select the PRODUCTS table in the SH schema and click Finish.

2. Open the Transforms category in the Component Palette, and add a Join node to the workflow.

3. Connect the SALES node to the Join node, and then connect the PRODUCTS node to the Join node. When done, the workflow should look similar to the following screenshot:
4. Double-click the Join node (or select Edit from the right-click menu) to open the Edit Join Node window.

5. On the Join tab, click the Add icon (+) to open the Edit Join Column dialog box.

![Edit Join Node Window]

6. In the Edit Join Column dialog box, do the following:
   a. Select SALES for Source 1 and PRODUCTS for Source 2.
   b. Select PROD_ID as the join column for both tables.
   c. Click the Add (v) button to generate the join condition.
   The dialog box should now look similar to the following screenshot:

![Edit Join Column Dialog]

d. Click OK to save the join condition.

7. Still in the Edit Join Node window, click the Columns tab, and then do the following:
   a. Deselect the Automatic Settings option (by default, Data Miner includes all columns from the joined tables).
   b. Select all of the columns in the PRODUCTS table except for PROD_NAME
   c. Click the Remove icon (red “X”).
   The Columns tab should now look similar to the following screenshot:
8. Click OK in the Edit Join Node window to complete the Join node definition.

9. Execute the join by selecting the Join node and clicking the Run icon ( ).

10. Select the Assoc Build node and copy it to the clipboard.

11. Paste the clipboard contents into the workflow. A second association node, named Assoc Build 1, appears in the workflow.

   Note: The new Assoc Build 1 node retains all of the definition specifications that you created previously in this lesson for the Assoc Build node.

12. Drag the Assoc Build 1 node to the right of the Join node.
13. Connect the Join node to the Assoc Build 1 node.

![Diagram showing the connection between nodes]

14. Right-click the Assoc Build 1 node and select Edit.

15. In the Edit Association Build Node dialog box, do the following:
   a. Change the Item ID option from PROD_ID to PROD_NAME.
   b. Click OK to save the definition.
16. Run the Assoc Build 1 node. When the build process is complete, the workflow should look similar to the following screenshot:

![Workflow Diagram](image)

17. Select **View Models > ASSOC_AP** from the Assoc Build 1 node’s right-click menu.

18. On the Rules tab of the model results, sort the output by the **Confidence** column, in **Descending** order.

19. Click the **Query** button to refresh the rules output, which should be similar to the following screenshot.

   **Note:** Your output results may be slightly different.

   ![Rules Output](image)

   **Note:** Now all of the antecedent and consequent values show product names, rather than the product ID values.

   Leave the ASSOC_AP window open, and proceed to practice 8-4.
Practice 8-4: Applying Model Rule Filters and Creating Model Details

Overview
In this practice, you define and apply filters to narrow the model rule output. You also create and run a Model Details node, and then view the results.

Assumptions
Practice 8-3 has been completed successfully.

Tasks
1. In the ASSOC_AP window open, click the More (v) button (just below the Query button). This action displays the filter options for the Rules tab.
   
   Note: The “More (v)” button text toggles to “Less (^).”

Notes
- Additional filter options appear in the top half of the window, although they are not enabled. By default, the User Filter option is not selected.
- The Rules pane is displayed in the lower half of the window.
2. Select the **Use Filter** option, and then click the **Add Filter** icon ( ), as shown here:

![Add Filter Icon]

3. In the Add Item Filters dialog box, do the following:
   a. Scroll down the product attribute list and select “**CD-R with Jewel Cases, Pack of 12**.”
   b. Ensure that the filter applies to rules using **Consequent** items (this option is in the lower right of the dialog box).
   c. Click **OK** to save the filter.

The Item Filters pane should now look similar to the following screenshot:

![Add Item Filters Dialog]

4. Click **Query** to refresh the Rules output with the new filter applied.

Your results will be similar to the output shown in the following screenshot:
Note: Now the rule output is narrowed down to 97 cases out of the 7,934.

5. Using the extra filter options on the left, change the Minimum Confidence (%) to 90 and re-query the data. Your results should be similar to those shown in the following screenshot:

Note: Now the rule output is narrowed down to 48 cases out of the 7,934.
6. Dismiss the ASSOC_AP window.

With an association rules model, you do not use an Apply node for deployment purposes. Instead, a Model Details node is used to generate output results that may be deployed.

7. To use a Model Details node, do the following:
   a. Add a Model Details node to the workflow from the Models category.
   b. Connect the Assoc Build 1 node to the Model Details node.
   c. Execute the Model Details node by selecting Run from the node’s right-click menu.

8. After the run process completes, select View Data from the node’s right-click menu.

9. As before, sort the results by the CONFIDENCE column in descending order.

Result: The Model Details output looks similar to that shown in the following screenshot:

10. Close the Model Details window, save the Market Basket Analysis workflow, and close the workflow window.
Practices for Lesson 9: Performing Anomaly Detection

Chapter 9
Practices for Lesson 9

Practices Overview
In these practices, you create a workflow that uses Anomaly Detection models to identify possible fraudulent automobile claims.
Practice 9-1: Creating an Anomaly Detection Model

Overview
In this practice, you create a new workflow, select the source data, and define an anomaly detection model for this case study.

Assumptions
The ABC Insurance project has been created.

Tasks
1. Create a new workflow in the exiting ABC Insurance project by doing the following:
   a. On the Data Miner tab, right-click the ABC Insurance project and select New Workflow.
   b. In the Create Workflow dialog box, enter Claims Anomaly Detection as the name and click OK.

   Result: The workflow environment should now look similar to the following screenshot:

   ![Workflow Environment Screenshot]

2. Using the component palette, drag and drop a Data Source node onto the workflow.

   Result: The Define Data Source Wizard opens automatically.

3. In Step 1 of the Define Data Source Wizard, select the DMUSER.CLAIMS table and click Finish.

   ![Define Data Source - Step 1 of 2 Screenshot]

   Select Table

<table>
<thead>
<tr>
<th>Select Column</th>
<th>Available Tables/Views:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Type</td>
</tr>
<tr>
<td>DMUSER.CLAIMS</td>
<td>TABLE</td>
</tr>
<tr>
<td>DMUSER.CUST_INSUR_LTV</td>
<td>TABLE</td>
</tr>
<tr>
<td>DMUSER.CUST_INSUR_LTV_APPLY</td>
<td>TABLE</td>
</tr>
</tbody>
</table>
4. Open the **Models** category in the Component Palette and drag the **Anomaly Detection** node from the palette to the workflow.

5. Connect the **CLAIMS** node to the **Anomaly Build** node.

6. After connecting the nodes, right-click the **Anomaly Build** node and select **Edit**.

7. In the **Edit Anomaly Detection Build Node** window, select **POLICYNUMBER** as the Case ID attribute.

8. Still in the **Edit Anomaly Detection Build Node** window, click the **Advanced** button.

9. In the Advanced Settings window, click the **Algorithm Settings** tab.

10. Change the Kernel Function option from System Determined to **Linear**.
11. Click **OK** in the Advance Settings window.

12. Click **OK** in the Edit Anomaly Detection Build Node window to save the model definition.

13. Build the anomaly detection model by selecting the **Anomaly Build** node and clicking the **Run** icon ( ).

**Result:** When the build process is complete, the workflow nodes should look similar to the following screenshot:
Practice 9-2: Viewing Model Results and Applying the Model

Overview
During an anomaly detection model build process, a training pass is made through the data to create a homogenous result set. Then a second pass is completed to flag records that do not fit the normal profile. These are the anomalous records.

In this practice, you:

- View the model build results
- Apply the model to the source data
- View the model apply results

Assumptions
Practice 9-1 has been successfully completed.

Tasks
1. View the model results: Right-click the Anomaly Build node, select View Models, and select the SVM model results.

   ![Diagram of Claims Anomaly Detection with Anomaly Build node]

   **Result:** By default, the output shows data for the Normal predictive class, as shown on the next page.

   **Note:** Your coefficient values may differ slightly from those shown in the following screenshot.
2. Click the Compare tab to see the propensity for normal and anomalous outcomes side-by-side.

3. Close the ANOM_SVM_#_# window by clicking the Close icon (X) on the tab.
4. In the Component Palette, open the “Evaluate and Apply” category, and drag an Apply node to the workflow, to the right of the Anomaly Build node.

5. Create the following two connections to the Apply node:
   - Connect the CLAIMS node to the Apply node.
   - Connect the Anomaly Build node to the Apply node.

   **Result:** The workflow should look similar to the following screenshot:

![Workflow Screenshot](image)

6. As you learned previously, you can remove or add columns to the model output by editing the Apply node. Double-click the Apply node (or select Edit from the node’s right-click menu).

   **Result:** The following columns for the SVM algorithm appear in the Output Apply Columns list:
   - Prediction
   - Prediction Probability

![Edit Apply Node](image)

7. In the Edit Apply Node dialog box, do the following:
   a. At the bottom of the Apply Columns tabbed pane, select **Apply Columns First** as the Default Column Order.
   b. Click the **Data Columns** tab, and click the **Edit Data Columns icon** ( ) to open the Edit Output Data Column dialog box.
   c. In the Edit Output Data Column dialog box, move all data columns to the selected list.
d. Click OK to close the Edit Output Data Column dialog box.
e. Click OK in the Edit Apply Node dialog box to save the model output format.

8. Run the Apply node by doing either of the following:
   • Select Run from the Apply node’s right-click menu.
   • Select the node and clicking the Run icon ( ).

Result: When the run process is complete, all nodes in the workflow contain green check marks in the node border.

9. After the Apply process is done, select View Data from the Apply node’s right-click menu.
10. As with all Apply data output, you can sort the results. Sort the output by the **Prediction** column (ascending order) and **Probability** column (descending order).

### Notes
- The records are sorted first by a prediction of “0” and are considered anomalous. Each record is then sorted by the probability of the anomalous finding, showing the most probable cases of fraud at the top of the output.
- Recall that anomaly detection models detect abnormal cases, but do not give reasons. Therefore, in terms of the model results, examination of the flagged records is necessary to determine why the records are detected as anomalous.
- Your results may be slightly different from those shown in the example below.
11. Resize the output columns in order to view more of that data, in a fashion similar to that shown in the following screenshot.

![Screenshot of data columns]

From our example, the following observations can be made:

- Some of the records that are predicted to have the highest probability of anomaly contain questionable data values when comparing the make and age of the wrecked vehicle with the reported value of that vehicle.

- In the suspicious records, highlighted above, the stated price of the vehicle seems too high for the make and age of the vehicle.

- The data for these records indicate the need for further investigation of these particular claims for possible fraud.

12. When you are done viewing the Apply results, close the Apply window, save your workflow, and close the workflow window.
Practices for Lesson 10: Deploying Data Mining Results

Chapter 10
Practices for Lesson 10

Practices Overview
In these practices, you deploy data mining results by using two methodologies: persisting results in a database table or view, and deploying SQL code.

You complete the following tasks:

• Create a database table to store the predictive results from the association rules model workflow that you created previously.
• Deploy SQL code from the classification model workflow that you created previously.
• View the deployed results by using SQL Developer.
Practice 10-1: Creating a Table for Data Mining Results

Overview
In this practice, you create a database table to store the market basket analysis predictive results that were created in your association rules workflow. You then view data in the table by using both Data Miner and SQL Developer.

Assumptions
Practices for Lesson 8 have been completed successfully.

Tasks
1. On the Data Miner tab, drill down on the Retail project.

2. Double-click the Market Basket Analysis workflow to open it in the workflow window.

3. Drag the Create Table or View option to the workflow.

Note: The node name is automatically generated. In this case: OUTPUT_1_11. Your OUTPUT table name may be different.

4. Connect the Model Details node to the OUTPUT node.

5. Right-click the OUTPUT node and select Edit.
6. In the “Edit Create Table or View Node” window, do the following:
   a. Uncheck the **Auto Input Columns Selection** option.
   b. Select all columns except for ANTECENDENT_ITEMS, CONFIDENCE, CONSEQUENT_ITEMS, ID, and LIFT.
   c. Click the **Delete selected rows** icon ( ). The selected columns should look similar to the following screenshot:
   d. Click **OK** to save the table definition.
7. Right-click the OUTPUT node and select Run.

**Result:** The table is created in the dmuser schema. When the process is complete, all workflow nodes contain a green check-mark in the border.

8. Right-click the OUTPUT node and select View Data. The table output looks something like this (your output may be slightly different):

<table>
<thead>
<tr>
<th>ID</th>
<th>ANTECEDENT_ITEMS</th>
<th>CONSEQUENT_ITEMS</th>
<th>CONFIDENCE</th>
<th>LIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>113 PROD_NAME,Comic Book Heroes = PROD_NAME,Martial Arts Champion...</td>
<td>0.734</td>
<td>18.1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>114 PROD_NAME,Martial Arts Champion...</td>
<td>PROD_NAME,Comic Book Heroes = PROD_NAME,Adventures with num...</td>
<td>0.5199</td>
<td>18.1</td>
</tr>
<tr>
<td>3</td>
<td>115 PROD_NAME,Comic Book Heroes = PROD_NAME,Adventures with num...</td>
<td>0.8433</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>116 PROD_NAME,Adventures with num...</td>
<td>PROD_NAME,Comic Book Heroes =</td>
<td>0.3318</td>
<td>11.3</td>
</tr>
<tr>
<td>5</td>
<td>117 PROD_NAME,Comic Book Heroes = PROD_NAME,Envoy External 6X CD...</td>
<td>0.4119</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>118 PROD_NAME,Envoy External 6X CD...</td>
<td>PROD_NAME,Comic Book Heroes =</td>
<td>0.1441</td>
<td>5.1</td>
</tr>
<tr>
<td>7</td>
<td>119 PROD_NAME,Comic Book Heroes = PROD_NAME,PCMCIA modem/fax 1...</td>
<td>0.4655</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>120 PROD_NAME,PCMCIA modem/fax 1...</td>
<td>PROD_NAME,Comic Book Heroes =</td>
<td>0.1059</td>
<td>3.7</td>
</tr>
<tr>
<td>9</td>
<td>121 PROD_NAME,Comic Book Heroes = PROD_NAME,Model K8822S Cordle...</td>
<td>0.364</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>122 PROD_NAME,Model K8822S Cordle...</td>
<td>PROD_NAME,Comic Book Heroes =</td>
<td>0.1127</td>
<td>4.0</td>
</tr>
<tr>
<td>11</td>
<td>123 PROD_NAME,Envoy 256MB - 40G3...</td>
<td>PROD_NAME,17” LCD w/built-in H...</td>
<td>0.4987</td>
<td>16.3</td>
</tr>
<tr>
<td>12</td>
<td>124 PROD_NAME,17” LCD w/built...</td>
<td>PROD_NAME,Envoy 256MB - 40G3...</td>
<td>0.457</td>
<td>16.3</td>
</tr>
<tr>
<td>13</td>
<td>125 PROD_NAME,Envoy 256MB - 40G3...</td>
<td>PROD_NAME,Mini DV Camcorder wi...</td>
<td>0.5511</td>
<td>18.2</td>
</tr>
<tr>
<td>14</td>
<td>126 PROD_NAME,Mini DV Camcorder wi...</td>
<td>PROD_NAME,Envoy 256MB - 40G3...</td>
<td>0.499</td>
<td>18.2</td>
</tr>
<tr>
<td>15</td>
<td>127 PROD_NAME,External 6X CD-ROM...</td>
<td>PROD_NAME,External 8X CD-ROM...</td>
<td>0.61</td>
<td>6.9</td>
</tr>
<tr>
<td>16</td>
<td>128 PROD_NAME,External 8X CD-ROM...</td>
<td>PROD_NAME,External 6X CD-ROM...</td>
<td>0.5232</td>
<td>6.9</td>
</tr>
<tr>
<td>17</td>
<td>129 PROD_NAME,External 6X CD-ROM...</td>
<td>PROD_NAME,Internal 6X CD-ROM =</td>
<td>0.3189</td>
<td>5.5</td>
</tr>
<tr>
<td>18</td>
<td>130 PROD_NAME,Internal 6X CD-ROM = PROD_NAME,External 6X CD-ROM...</td>
<td>0.419</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>131 PROD_NAME,External 6X CD-ROM...</td>
<td>PROD_NAME,Internal 8X CD-ROM =</td>
<td>0.3233</td>
<td>4.3</td>
</tr>
<tr>
<td>20</td>
<td>132 PROD_NAME,Internal 8X CD-ROM = PROD_NAME,External 6X CD-ROM...</td>
<td>0.3322</td>
<td>4.3</td>
<td></td>
</tr>
</tbody>
</table>

9. Close the OUTPUT window.

Next, you view the table results in a SQL Developer worksheet window.

10. Using the SQL Developer main menu, select View > Connections.
11. On the SQL Developer Connections tab, drill down on `dmuser > Tables` to view all associated database table objects sorted in alphabetical order.

![Connections tree](image)

12. Scroll down in the list of tables until you see the OUTPUT table that you just created. Then click the table name (your OUTPUT table name may be different).

**Result:** The SQL Developer multi-tabbed table window opens. The Columns tab is displayed by default.

![Table window](image)

13. Click the **Data** tab to view the predictive results, which should look similar to the following screenshot:

![Predictive results](image)
14. Click the Sort button, and sort the data by CONFIDENCE, in descending order. Then click OK.

![Sort Columns](image)

**Result:** The output shows records sorted by Confidence values.

<table>
<thead>
<tr>
<th>ID</th>
<th>ANTECEDENT_ITEMS</th>
<th>CONSEQUENT_ITEMS</th>
<th>CONFIDENCE</th>
<th>LIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2760 PROD_NAME,Musick CD-R , PROD_NAME,Musick CD-R</td>
<td>PROD_NAME,CD-R with Jewel Cases</td>
<td>0.95710407239</td>
<td>9.2023</td>
</tr>
<tr>
<td>2</td>
<td>5121 PROD_NAME,CD-R, High Speed, Pack of 10</td>
<td>PROD_NAME,CD-R with Jewel Cases</td>
<td>0.99954164057</td>
<td>9.2471</td>
</tr>
<tr>
<td>3</td>
<td>4752 PROD_NAME,Drumkit CD-R, Pack of 10 =</td>
<td>PROD_NAME,CD-R with Jewel Cases</td>
<td>0.95280349143</td>
<td>9.2209</td>
</tr>
<tr>
<td>4</td>
<td>6161 PROD_NAME,External 6X CD-ROM =</td>
<td>PROD_NAME,6/8 Documentation Set - English =</td>
<td>0.95105801045</td>
<td>6.2748</td>
</tr>
<tr>
<td>6</td>
<td>5928 PROD_NAME,2500MB Memory Card =</td>
<td>PROD_NAME,Microsoft 128MB Memory Card =</td>
<td>0.9540733247</td>
<td>10.5526</td>
</tr>
<tr>
<td>7</td>
<td>572 PROD_NAME,Laptop Discs, 4.7GB, Pack of S</td>
<td>PROD_NAME,Model S80D273 Black Ink Cartridge =</td>
<td>0.954005301</td>
<td>10.6704</td>
</tr>
<tr>
<td>8</td>
<td>5172 PROD_NAME,External 6X CD-ROM =</td>
<td>PROD_NAME,6/8 Documentation Set - English =</td>
<td>0.9476940202</td>
<td>6.2524</td>
</tr>
<tr>
<td>9</td>
<td>7699 PROD_NAME,1/2quot; Sata cd drive, Pack of</td>
<td>PROD_NAME,Model S80D273 Black Ink Cartridge =</td>
<td>0.9476433295</td>
<td>10.6623</td>
</tr>
</tbody>
</table>

15. In the Filter field, enter the following WHERE clause, and then press Enter:

   CONFIDENCE>.93 and LIFT>13

**Result:** The sorted display now includes only six cases.

![Filter Field](image)

**Note:** This table may be queried by any SQL-based application or tool.

16. Close the SQL Developer OUTPUT window.

17. Close the SQL worksheet window with the title dmuser.

18. On the SQL Developer Connections tab, collapse the dmuser node, but leave the Connections tab open.
19. Click the **Data Miner** tab.

20. Close the Market Basket Analysis workflow window.
Practice 10-2: Deploying the SQL Code for Data Mining Results

Overview
In this practice, you perform the following tasks:

- Open the Buy Insurance workflow, which uses a classification model.
- Deploy the SQL code for the model's predictive results.
- Execute the SQL code in SQL Developer.

Assumptions
Practices for Lesson 5 have been completed successfully.

Tasks
1. On the Data Miner tab, open the ABC Insurance project, and double-click the Buy Insurance workflow to open it in the workflow window.

2. Right-click the Apply node and select Deploy > SQL to Clipboard.
3. In the Deploy Code information dialog box, click **OK** to continue.

![Deploy Code dialog box](image)

The deployed SQL includes SQL generated by the current node and all of its parent nodes that are data providers. The SQL lineage ends when it encounters a node that represents persisted objects, such as tables or models.

- **Skip This Message Next Time**

![Screenshot of Deploy Code](image)

4. You now open a SQL Developer worksheet, paste the Data Miner code, and execute the SQL to view results in SQL Developer. Perform the following actions:

a. Click the SQL Developer **Connections** tab.

b. Right-click the **dmuser** connection and select **Open SQL Worksheet**.

![Connections tab](image)

Result: A SQL Developer worksheet window with the name dmuser opens.
5. Right-click inside the **dmuser worksheet window** and select **Paste**.

**Result:** The deployed data mining code appears in the SQL worksheet.

**Note:** The “N$” table names in your code will probably be different.

```sql
/* SQL Deployed by Oracle SQL Developer 3.1.07.42 from Node "Apply", Workflow "Buy Insurance" */

--ALTER SESSION FOR OPTIMIZER
ALTER SESSION set "optimizer_reuse_cost_annotations"='false';
ALTER SESSION set NLS_NUMERIC_CHARACTERS='"","';

WITH
/* Start of sql for node: INSUR_CUST_LTV_SCORE */
N$10013 as (select /*+ inline */ "INSUR_CUST_LTV_APPLY"."MARITAL_STATUS",
"INSUR_CUST_LTV_APPLY"."STATE",
"INSUR_CUST_LTV_APPLY"."CREDIT_BALANCE",
"INSUR_CUST_LTV_APPLY"."TIME_AS_CUSTOMER",
"INSUR_CUST_LTV_APPLY"."CUSTOMER_ID",
"INSUR_CUST_LTV_APPLY"."MORTGAGE_AMOUNT",
"INSUR_CUST_LTV_APPLY"."BANK_FUNDS",
"INSUR_CUST_LTV_APPLY"."N_OF_DEPENDENTS",
"INSUR_CUST_LTV_APPLY"."HAS_CHILDREN",
"INSUR_CUST_LTV_APPLY"."SALARY",
"INSUR_CUST_LTV_APPLY"."SEX",
"INSUR_CUST_LTV_APPLY"."PROFESSION",
"INSUR_CUST_LTV_APPLY"."CREDIT_CARD_LIMITS",
"INSUR_CUST_LTV_APPLY"."REGION",
"INSUR_CUST_LTV_APPLY"."HOUSE_OWNERSHIP",
"INSUR_CUST_LTV_APPLY"."N_TRANS_WEB_BANK",
"INSUR_CUST_LTV_APPLY"."BUY_INSURANCE",
"INSUR_CUST_LTV_APPLY"."MONTHLY_CHECKS_WRITTEN",
"INSUR_CUST_LTV_APPLY"."N_TRANS_KIOSK",
"INSUR_CUST_LTV_APPLY"."AGE",
"INSUR_CUST_LTV_APPLY"."MONEY_MONTHLY_OVERDRAWN",
"INSUR_CUST_LTV_APPLY"."LTV",
"INSUR_CUST_LTV_APPLY"."T_AMOUNT_AUTOM_PAYMENTS",
"INSUR_CUST_LTV_APPLY"."N_TRANS_TELLER",
"INSUR_CUST_LTV_APPLY"."CHECKING_AMOUNT",
"INSUR_CUST_LTV_APPLY"."N_TRANS_ATM",
"INSUR_CUST_LTV_APPLY"."LTV_BIN",
```
6. Click the **Run Statement** icon ( Run Statement ) to execute the SQL code.

**Result:** A Query Result window opens with the results.

### Notes

- The “N$” table names and column names for your model prediction output will be different than those shown above.

- In addition to the customer ID, there are three output columns. In our example, they are: the prediction (CLAS_DT_1_1_PRED), the probability (CLAS_DT_1_1_PROB), and the cost (CLAS_DT_1_1_PCST).

- The prediction and probability of the prediction are straightforward. But what is the cost?
  - Recall that each algorithm in a build node has certain advanced settings that may be modified. These settings are organized into Data Usage, Algorithm Settings, and Performance Settings tabs.
  - All advanced setting options for each algorithm are given a default setting. The default Performance Settings option for a decision tree model is Balanced, as shown in the following screenshot of the Advanced Settings dialog box of the Class Build node in your workflow.
− In order to achieve a balanced model build, a decision tree model is provided a cost matrix.
− This matrix can be seen on the Cost Matrix/ Benefit tab of the DT model viewer, as shown below.

(To view this setting, right-click the successfully built Class Build node and select: View Model > CLAS_DT_#_# > Settings tab > Cost Matrix/Benefit tab.)

− Therefore, when the Balanced option is used for a decision tree model, the prediction of YES/NO is actually driven by the lowest cost value.
− For each case in the data, the ultimate prediction is YES if the cost for YES is lower than the cost for NO.

Using this information, in Practice 10-3 you will perform the following tasks:
   a. Modify the Class Build node in your workflow to use a Decision Tree algorithm that uses the Natural performance setting, rather than Balanced.
   b. Rerun the Apply node.
   c. Deploy the new code and view the differences in the model prediction.
Practice 10-3: Modifying Classification Models and Redeploying the Code

Overview
In this practice, you modify the advanced settings of the decision tree model to use a different algorithm option. Then you rebuild the model and deploy the SQL code to examine the results.

Assumptions
Practice 10-2 has been completed successfully.

Tasks
1. In the workflow, right-click the Class Build node and select Edit.

2. In the Edit Classification Build Node window, select the Decision Tree algorithm, and click the Duplicate Selected Model icon ( ).

---

Copyright © 2012, Oracle and/or its affiliates. All rights reserved.
Practices for Lesson 10: Deploying Data Mining Results
Chapter 10 - Page 14
Result: A new Decision Tree algorithm is added to the list (CLAS_DT_2_1, in our example).

<table>
<thead>
<tr>
<th>Name</th>
<th>Algorithm</th>
<th>Date</th>
<th>Auto</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAS_GLM_1_1</td>
<td>Generalized Linear Model</td>
<td>2/15/12 1:44 AM</td>
<td></td>
</tr>
<tr>
<td>CLAS_SVM_1_1</td>
<td>Support Vector Machine</td>
<td>2/15/12 1:44 AM</td>
<td></td>
</tr>
<tr>
<td>CLAS_DT_1_1</td>
<td>Decision Tree</td>
<td>2/15/12 1:44 AM</td>
<td></td>
</tr>
<tr>
<td>CLAS_NB_1_1</td>
<td>Naive Bayes</td>
<td>2/15/12 1:44 AM</td>
<td></td>
</tr>
<tr>
<td>CLAS_DT_2_1</td>
<td>Decision Tree</td>
<td>2/15/12 1:44 AM</td>
<td></td>
</tr>
</tbody>
</table>

3. With the new DT model selected, click the Advanced button.

4. On the Performance Settings tab of the Advance Settings window, select the Natural option.

![Advanced Settings window]

Note: This option instructs Oracle Data Miner to let the model achieve maximum overall accuracy, rather than maximum average accuracy. Therefore, there is no need to include a cost matrix to weight the target class values to balance them out.

5. Click OK in the Advanced Settings window. Then click OK in the Edit Classification Build Node window to save the change.

Result: The Property Inspector of the Class Build node now shows five model setting options, including the new Decision Tree algorithm that has not yet been built.

![Class Build Property Inspector]

---

Copyright © 2012, Oracle and/or its affiliates. All rights reserved.

Practices for Lesson 10: Deploying Data Mining Results
Chapter 10 - Page 15
6. Right-click the **Apply Model** node and select **Edit** from the menu.

In the Edit Apply Node window, you can view the new settings for the updated DT algorithm.

<table>
<thead>
<tr>
<th>Column</th>
<th>Function</th>
<th>Parameter(s)</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLAS_DT_1_1_PRED</td>
<td>Prediction</td>
<td></td>
<td>CLAS_DT_1_1</td>
</tr>
<tr>
<td>CLAS_DT_1_1_PROB</td>
<td>Prediction Probability</td>
<td>Target: &lt;Most Likely&gt;</td>
<td>CLAS_DT_1_1</td>
</tr>
<tr>
<td>CLAS_DT_1_1_PCST</td>
<td>Prediction Cost</td>
<td>Target: &lt;Most Likely&gt;</td>
<td>CLAS_DT_1_1</td>
</tr>
<tr>
<td>CLAS_DT_2_1_PRED</td>
<td>Prediction</td>
<td></td>
<td>CLAS_DT_2_1</td>
</tr>
<tr>
<td>CLAS_DT_2_1_PROB</td>
<td>Prediction Probability</td>
<td>Target: &lt;Most Likely&gt;</td>
<td>CLAS_DT_2_1</td>
</tr>
</tbody>
</table>

**Notes**
- The new DT model (CLAS_DT_2_1 in our example) contains only two output columns, as opposed to three columns for the original DT model.
- The new DT model does not have a Prediction Cost column, because this model will not use the cost matrix due to the performance setting selection of Natural.

7. Click **OK** to close the Edit Apply Node window.

8. Right-click the **Apply Model** node and select **Run**.

**Notes**
- This action automatically forces the Class Build node to run in order to build and test the new model, as well as run the Apply Model node.
- When the run process completes, the Property Inspector of the Class Build node should similar to the following screenshot:

<table>
<thead>
<tr>
<th>Model Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>CLAS_GLN_1_1</td>
</tr>
<tr>
<td>CLAS_SVM_1_1</td>
</tr>
<tr>
<td>CLAS_DT_1_1</td>
</tr>
<tr>
<td>CLAS_DT_2_1</td>
</tr>
<tr>
<td>CLAS_NB_1_1</td>
</tr>
</tbody>
</table>

9. Right-click the **Apply Model** node and select **Deploy > Copy SQL to Clipboard**. Click **OK** in the Deploy Code message dialog box.
10. As you did in the previous practice:
   a. Click the SQL Developer Connections tab.
   b. Right-click the dmuser connection and select Open SQL Worksheet.

Result: A second worksheet window (dmuser~1) opens, like this:

11. Press Ctrl + V to paste the deployed code into the worksheet. The deployed code now shows five prediction columns: three for the first DT model, and two for the second DT model.

```sql
"INSUR_CUST_LTV_apply"."AGE",
"INSUR_CUST_LTV_apply"."AMOUNT_MONEY_OVERDRAWN",
"INSUR_CUST_LTV_apply"."LTV",
"INSUR_CUST_LTV_apply"."T_AMT_AMT_AUTOM_PAYMENTS",
"INSUR_CUST_LTV_apply"."N_TRANS_TELLER",
"INSUR_CUST_LTV_apply"."CHECKING_Amount",
"INSUR_CUST_LTV_apply"."N_TRANS_ATM",
"INSUR_CUST_LTV_apply"."LTV_DIN",
"INSUR_CUST_LTV_apply"."LAST",
"INSUR_CUST_LTV_apply"."N_MORTGAGES",
"INSUR_CUST_LTV_apply"."CAR_OWNERHIPS",
"INSUR_CUST_LTV_apply"."FIRST"
from "DMUSER"."INSUR_CUST_LTV_apply" ) )
/* End of sql for node: INSUR_CUST_LTV_SCORE */

/* Start of sql for node: Apply */
"N$10014" as (SELECT /*+ inline */
"CUSTOMER_ID"
, PREDICTION("DMUSER"."CLAS_DT_1_1" COST MODEL USING *) "CLAS_DT_1_1_PRED"
 , PREDICTION_PROBABILITY("DMUSER"."CLAS_DT_1_1" USING *) "CLAS_DT_1_1_PROB"
, PREDICTION_COST("DMUSER"."CLAS_DT_1_1" COST MODEL USING *) "CLAS_DT_1_1_PCST"
 , PREDICTION("DMUSER"."CLAS_DT_2_1" USING *) "CLAS_DT_2_1_PRED"
 , PREDICTION_PROBABILITY("DMUSER"."CLAS_DT_2_1" USING *) "CLAS_DT_2_1_PROB"
from "N$10013"
)
/* End of sql for node: Apply */

select * from "N$10014";
```
12. Click the **Run Statement** icon ( ) to execute the SQL code.

The resulting should look similar to the following screenshot:

![Screenshot of SQL execution result]

**Notes**

- In the output shown above, the prediction outcomes (Yes or No) for the two different DT models seem to be quite similar, although the prediction probabilities are not exactly the same.

- However, for customer CU13551 (row 5 above), there is a difference in the predictions between the two DT models.
  - The first DT model, using a Balanced performance setting (employing the cost matrix), predicts that the customer will buy insurance with a probability of 55%.
  - The second DT model, using a Natural performance setting, predicts that the customer will not buy insurance with a probability of 98%.

- Other classification models (using GLM, SVM, or NB algorithms) do not use cost to get balanced performance, so they do not generate any cost output.

- However, if you were to use the “Tune options” for any of the models, you would be pushing a “score” time cost matrix into the model, which would generate costs for any model that was tuned.

13. Close the two dmuser worksheet windows, and close the High Value Customers workflow window.