Hands on Exercise for Data Mining – Part 1

Log on to elab using <u>your eID credentials</u>

Learn by Doing – Defining a Data Mining Model (From previous edition of Larson's Delivering BI book)

Business Need: The Maximum Miniatures Marketing Department wants to do a targeted mailing promoting the Mythic World line of figurines. Previous research has shown that the vast majority of customers who buy the Mythic World product line do not have any children living at home. Unfortunately, the list of addresses and demographics purchased for the Mythic World mailing does not include the number of children living at home. However, it does include the following facts about each household:

- Number of cars owned
- Marital status
- Whether the address is a rental property or occupied by the owner

The marketing department would like to find a way, using the three facts included in the mailing list data and the information known about current customers, to predict which households have no children living at home. The mailing will then be sent only to those households who are most likely to have no children living at home.

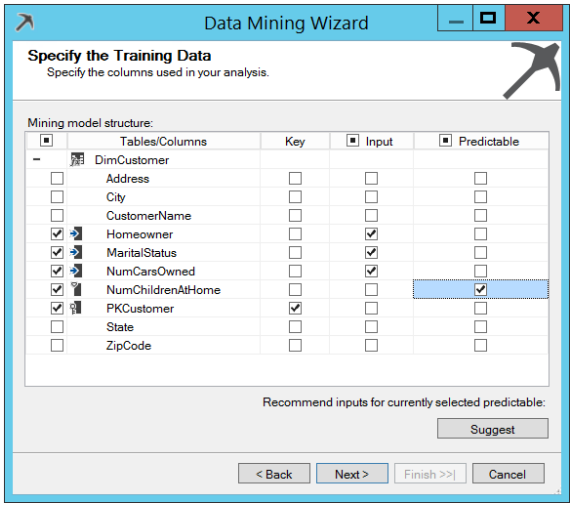
Steps to create data source for the mining models

1. Open Visual Studio 2017.
2. Click File→Open→Analysis Services Database.
   In the Connect To Database dialog box, enter **buscissql\cisbi** in the Server textbox and select **your analysis services database (e.g., if your last name is Smith, your analysis services database will be called SmithASD)** from the Database drop-down list. Click on the Browse button and select the folder where you would like the solution file to be stored (this can be your class folder for CIS570 (i.e., "S: drive")). Click OK to connect to the database.
3. Right-click the Data-Sources folder in the Solution Explorer window and select New Data Source from the context menu. The Data Source Wizard Appears.
4. Click Next. The Select how to define the connection page appears.
5. Click New. The Connection Manager dialog box appears. Enter **buscissql\cisbi** for Server name. Select **MaxMinMiningDM** from the Select or enter a database name drop-down list. Click OK.
6. Click Next in the Select how to define a connection dialog box.
7. Enter your **eID (e.g., if your eID is janesmith, enter colostate\janesmith) in the User name text box and your eID password in the Password text box**. Click Next. The Completing the Wizard page appears.
8. Enter **Max Min Mining DM** for Data source name, if it is not already there. Click Finish.
9. Right-click the Data Source Views folder in the Solution Explorer window and select New Data Source View from the context menu. The Data Source View Wizard appears.
10. Click Next. The Select a Data Source page appears.
11. Select the **Max Min Mining DM** data source that you just created. Click Next. The Select Tables and Views page appears.
12. Move the **DimCustomer** table to the Included objects list. Click Next. The Completing the Wizard page appears.

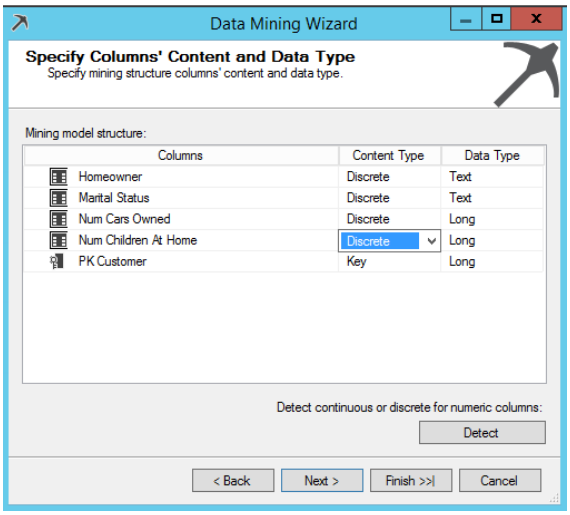
13. Enter **Max Min Mining DM** for Name, if it is not already there. Click Finish. The Data Source View Design tab appears.

Steps to Create a Mining Structure with One Mining Model

1. Right-click the Mining Structures folder in the Solution Explorer window and select New Mining Structure from the context menu. The Data Mining Wizard appears.
2. Click Next on the Welcome page. The Select the Definition Method page of the wizard appears.
3. Select From existing relational database or data warehouse radio button (it should be selected by default). Click Next. The Create the Data Mining Structure page of the wizard appears.
4. Select **Microsoft Decision Trees** (it may already be selected by default) from the drop-down list. Click Next. The Select Data Source View page of the wizard appears.
5. Select **Max Min Mining DM** from the Available data source views list box. Click Next. The Specify Table Types page of the wizard appears.
6. Check (if it is not already checked) the Case checkbox for the DimCustomer table. Click Next. The Specify the Training Data page of the wizard appears.
7. Check the Input checkbox for the following columns: **Homeowner**, **MaritalStatus**, and **NumCarsOwned**. Check the Predictable checkbox for the **NumChildrenAtHome** column. The Key checkbox should already be checked for **PKCustomer**. The Specify the Training Data page as shown below:

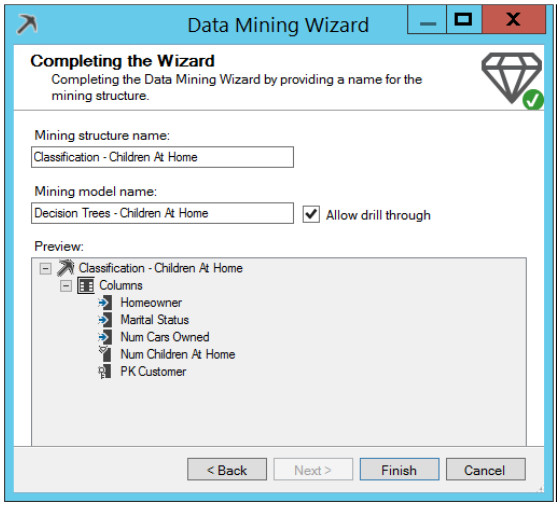


8. Click Next. The Specify Columns' Content and Data Type page of the wizard appears.
9. Change the Content Type for Num Cars Owned and Num Children At Home to Discrete. The Specify Columns' Content and Data Type page appears as the shown below.



Note: Even though the Num Cars Owned and Num Children At Home have Long integer data types, we know, in reality, that they have only a few discrete values. Therefore, we can change their Content Type to Discrete.

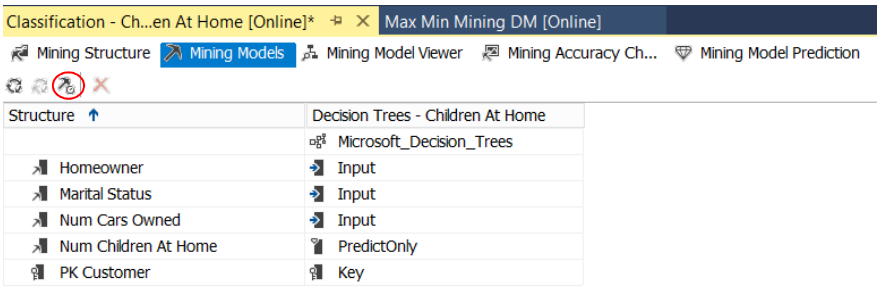
10. Click Next. The Create Testing Set page of the Wizard appears. Leave the default set at 30% for testing.
11. Click Next. The Completing the Wizard page appears. Enter **Classification – Children At Home** for Mining structure name. Enter **Decision Trees – Children At Home** for Mining Model name. Check Allow drill through. The Completing the Wizard page appears as shown below.



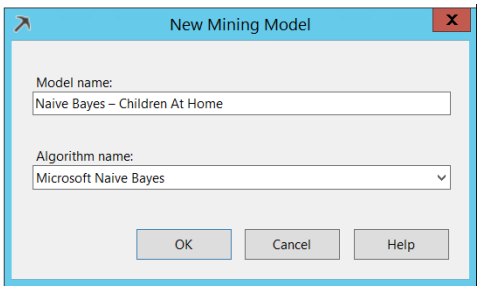
12. Click Finish. The data mining structure containing our new data mining model is created.

Steps to Add Mining Models to the Mining Structure

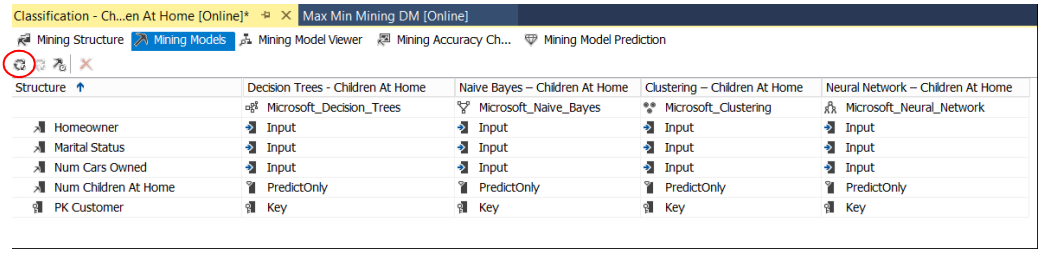
1. Select the Mining Models tab on the Data Mining Design tab (see below)



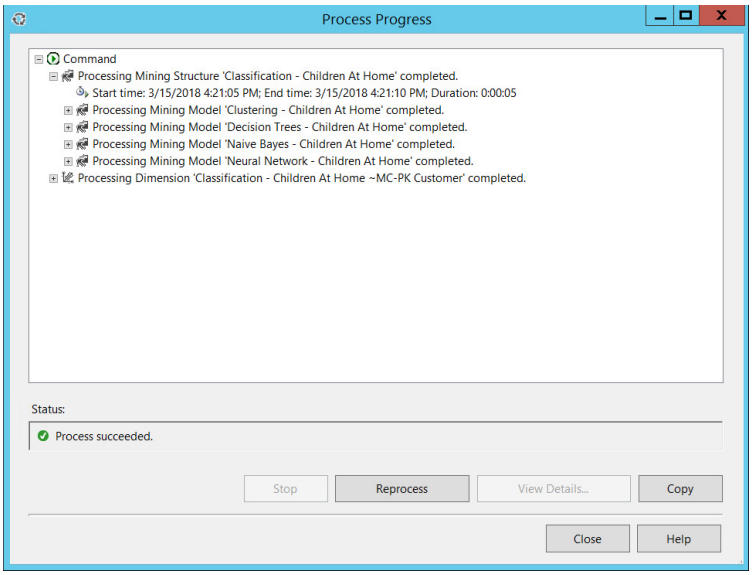
2. Click the Create a related mining model button on the Mining Models tab toolbar (circled above). The New Mining Model dialog box appears.
3. Enter **Naive Bayes – Children At Home** for Model name. Select **Microsoft Naive Bayes** from the Algorithm name drop-down list. The New Mining Model dialog box appears as shown below.



4. Click OK.
5. Again, click the Create a related mining model button on the Mining Models tab toolbar. The New Mining Model dialog box appears.
6. Enter **Clustering – Children At Home** for Model name. Select **Microsoft Clustering** from the Algorithm name drop-down list. Click OK.
7. Once more, click the Create a related mining model button on the Mining Models tab toolbar. The New Mining Model dialog box appears.
8. Enter **Neural Network – Children At Home** for Model name. Select **Microsoft Neural Network** from the Algorithm name drop-down list. Click OK. The Mining Models tab appears as shown below.



9. Click the Save All button on the toolbar. If a dialog box appears, click Save.
10. Click the Process the mining structure and all its related models button on the Mining Models tab toolbar (circled above). The Process Mining Structure dialog box appears. Click Run. The Process Progress dialog box appears.
11. When the models have all been trained and the process has been completed successfully, the Process Progress dialog box appears as shown below.



12. Click Close to exit the Process Progress dialog box.
13. Click Close to exit the Process Mining Structure dialog box.
14. Exit VS 2017.

We will discuss how to view and interpret the results in class.