* [Predicting Loan Defaulters (Bayesian Network)](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/clementine/example_bayesnet.htm)

# Predicting Loan Defaulters (Bayesian Network)

Bayesian networks enable you to build a probability model by combining observed and recorded evidence with "common-sense" real-world knowledge to establish the likelihood of occurrences by using seemingly unlinked attributes.

This example uses the stream named bayes\_bankloan.str, which references the data file named bankloan.sav. These files are available from the Demos directory of any IBM® SPSS® Modeler installation and can be accessed from the IBM SPSS Modeler program group on the Windows Start menu. The bayes\_bankloan.str file is in the streams directory.

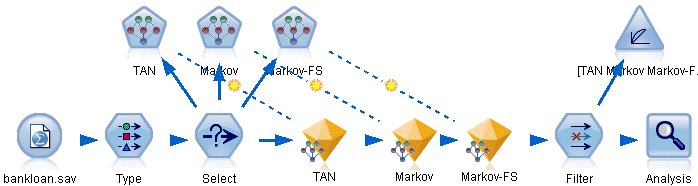
For example, suppose a bank is concerned about the potential for loans not to be repaid. If previous loan default data can be used to predict which potential customers are liable to have problems repaying loans, these "bad risk" customers can either be declined a loan or offered alternative products.

This example focuses on using existing loan default data to predict potential future defaulters, and looks at three different Bayesian network model types to establish which is better at predicting in this situation

**Building the Stream**

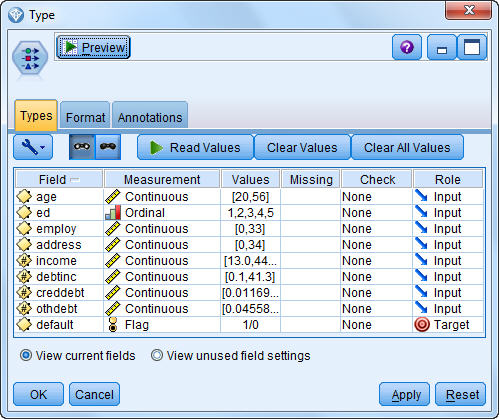
1. Add a Statistics File source node pointing to *bankloan.sav* in the *Demos* folder.

*Figure 1. Bayesian Network sample stream*



1. Add a Type node to the source node and set the role of the **default** field to **Target**. All other fields should have their role set to **Input**.
2. Click the **Read Values** button to populate the *Values* column.

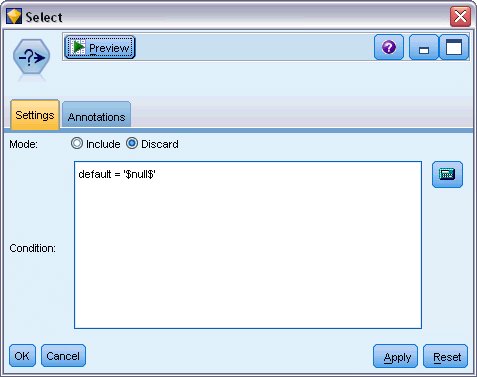
*Figure 2. Selecting the target field*



Cases where the target has a null value are of no use when building the model. You can exclude those cases to prevent them from being used in model evaluation.

1. Add a Select node to the Type node.
2. For Mode, select **Discard**.
3. In the Condition box, enter **default = '$null$'**.

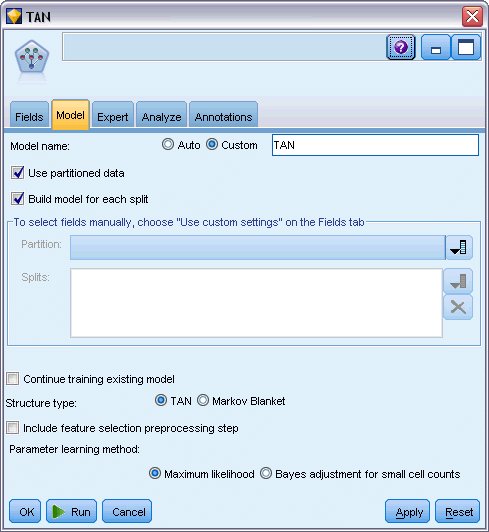
*Figure 3. Discarding null targets*



Because you can build several different types of Bayesian networks, it is worth comparing several to see which model provides the best predictions. The first one to create is a Tree Augmented Naïve Bayes (TAN) model.

1. Attach a Bayesian Network node to the Select node.
2. On the Model tab, for Model name, select **Custom** and enter TAN in the text box.
3. For Structure type, select **TAN** and click **OK**.

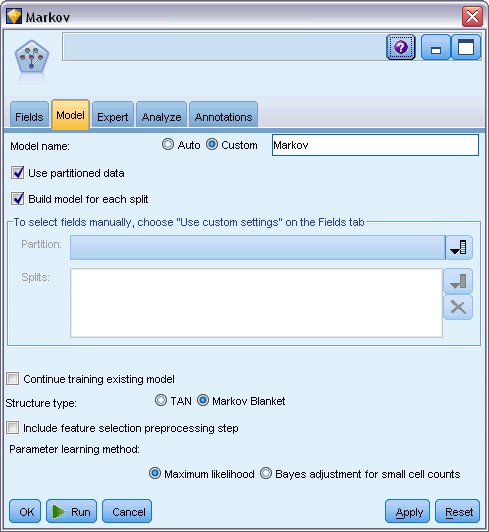
*Figure 4. Creating a Tree Augmented Naïve Bayes model*



The second model type to build has a Markov Blanket structure.

1. Attach a second Bayesian Network node to the Select node.
2. On the Model tab, for Model name, select **Custom** and enter Markov in the text box.
3. For Structure type, select **Markov Blanket** and click **OK**.

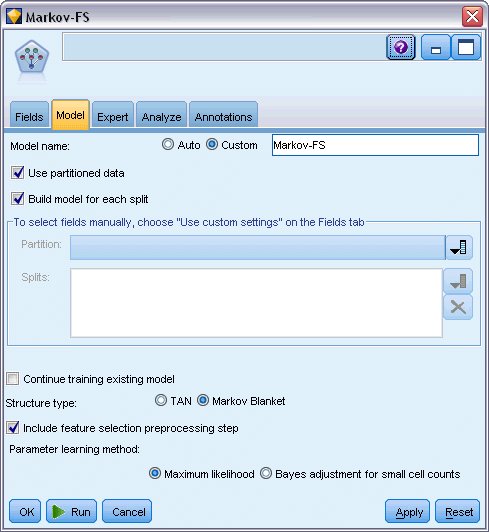
*Figure 5. Creating a Markov Blanket model*



The third model type to build has a Markov Blanket structure and also uses feature selection preprocessing to select the inputs that are significantly related to the target variable.

1. Attach a third Bayesian Network node to the Select node.
2. On the Model tab, for Model name, select **Custom** and enter Markov-FS in the text box.
3. For Structure type, select **Markov Blanket**.
4. Select **Include feature selection preprocessing step** and click **OK**.

*Figure 6. Creating a Markov Blanket model with Feature Selection preprocessing*



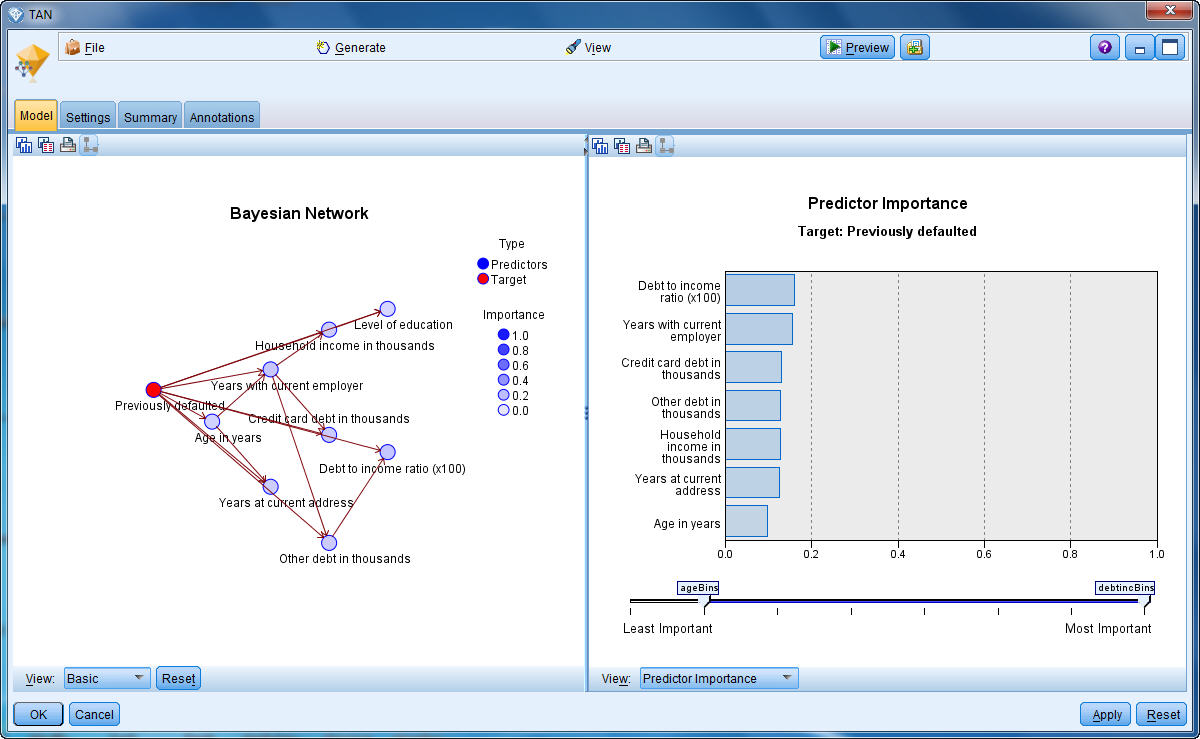
**Browsing the Model**

1. Run the stream to create the model nuggets, which are added to the stream and to the Models palette in the upper-right corner. To view their details, double-click on any of the model nuggets in the stream.

The model nugget Model tab is split into two panes. The left pane contains a network graph of nodes that displays the relationship between the target and its most important predictors, as well as the relationship between the predictors.

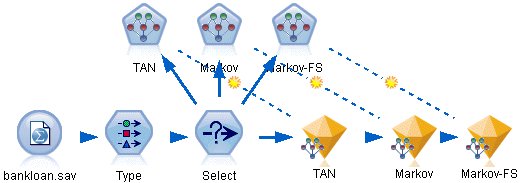
The right pane shows either *Predictor Importance*, which indicates the relative importance of each predictor in estimating the model, or *Conditional Probabilities*, which contains the conditional probability value for each node value and each combination of values in its parent nodes.

*Figure 1. Viewing a Tree Augmented Naïve Bayes model*



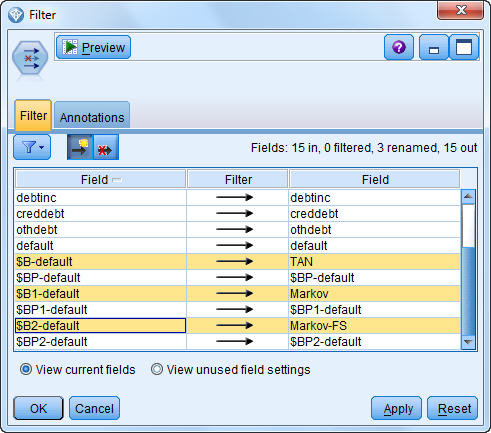
1. Connect the TAN model nugget to the Markov nugget (choose **Replace** on the warning dialog).
2. Connect the Markov nugget to the Markov-FS nugget (choose **Replace** on the warning dialog).
3. Align the three nuggets with the Select node for ease of viewing.

*Figure 2. Aligning the nuggets in the stream*



1. To rename the model outputs for clarity on the Evaluation graph that you'll be creating, attach a Filter node to the Markov-FS model nugget.
2. In the right *Field* column, rename $B-default as TAN, $B1-default as Markov, and $B2-default as Markov-FS.

*Figure 3. Rename model field names*

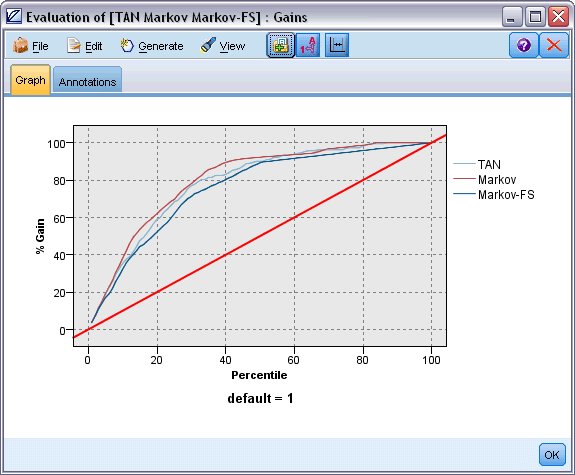


To compare the models' predicted accuracy, you can build a gains chart.

1. Attach an Evaluation graph node to the Filter node and execute the graph node using its default settings.

The graph shows that each model type produces similar results; however, the Markov model is slightly better.

*Figure 4. Evaluating model accuracy*

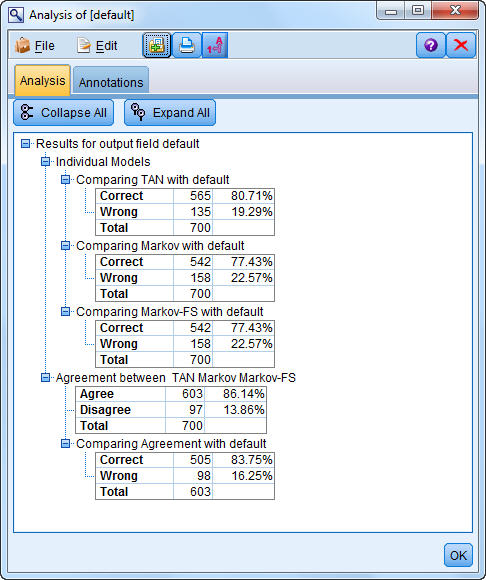


To check how well each model predicts, you could use an Analysis node instead of the Evaluation graph. This shows the accuracy in terms of percentage for both correct and incorrect predictions.

1. Attach an Analysis node to the Filter node and execute the Analysis node using its default settings.

As with the Evaluation graph, this shows that the Markov model is slightly better at predicting correctly; however, the Markov-FS model is only a few percentage points behind the Markov model. This may mean it would be better to use the Markov-FS model since it uses fewer inputs to calculate its results, thereby saving on data collection and entry time and processing time.

*Figure 5. Analyzing model accuracy*



Explanations of the mathematical foundations of the modeling methods used in IBM® SPSS® Modeler are listed in the *IBM SPSS Modeler Algorithms Guide*, available from the*\Documentation* directory of the installation disk.

Note also that these results are based on the training data only. To assess how well the model generalizes to other data in the real world, you would use a Partition node to hold out a subset of records for purposes of testing and validation.