* [Retraining a Model on a Monthly Basis (Bayesian Network)](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/clementine/example_bayesnet_retrain.htm)

# Retraining a Model on a Monthly Basis (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\_churn\_retrain.str, which references the data files named telco\_Jan.sav and telco\_Feb.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\_churn\_retrain.str file is in the streams directory.

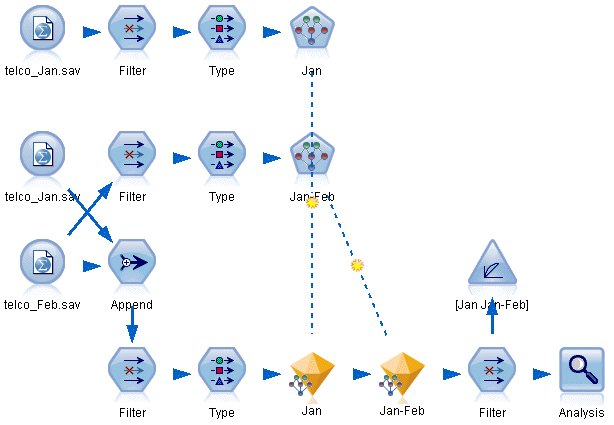
For example, suppose that a telecommunications provider is concerned about the number of customers it is losing to competitors (churn). If historic customer data can be used to predict which customers are more likely to churn in the future, these customers can be targeted with incentives or other offers to discourage them from transferring to another service provider.

This example focuses on using an existing month's churn data to predict which customers may be likely to churn in the future and then adding the following month's data to refine and retrain the model.

**Building the Stream**

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

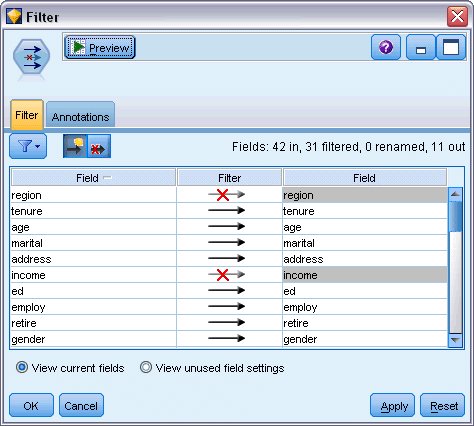
*Figure 1. Bayesian Network sample stream*



Previous analysis has shown you that several data fields are of little importance when predicting churn. These fields can be filtered from your data set to increase the speed of processing when you are building and scoring models.

1. Add a Filter node to the Source node.
2. Exclude all fields except *address*, *age*, *churn*, *custcat*, *ed*, *employ*, *gender*, *marital*, *reside*, *retire*, and *tenure*.
3. Click **OK**.

*Figure 2. Filtering unnecessary fields*



1. Add a Type node to the Filter node.
2. Open the Type node and click the **Read Values** button to populate the *Values* column.
3. In order that the Evaluation node can assess which value is true and which is false, set the measurement level for the *churn* field to **Flag**, and set its role to **Target**. Click **OK**.

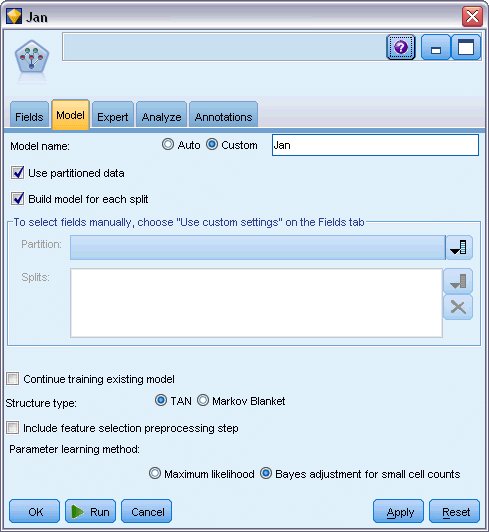
*Figure 3. Selecting the target field*



You can build several different types of Bayesian networks; however, for this example you are going to build a Tree Augmented Naïve Bayes (TAN) model. This creates a large network and ensures that you have included all possible links between data variables, thereby building a robust initial model.

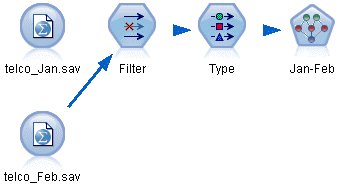
1. Attach a Bayesian Network node to the Type node.
2. On the Model tab, for Model name, select **Custom** and enter Jan in the text box.
3. For Parameter learning method, select **Bayes adjustment for small cell counts**.
4. Click **Run**. The model nugget is added to the stream, and also to the Models palette in the upper-right corner.

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



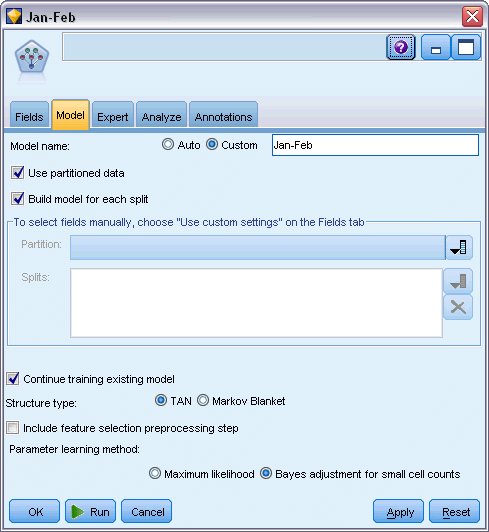
1. Add a Statistics File source node pointing to *telco\_Feb.sav* in the *Demos* folder.
2. Attach this new source node to the Filter node (on the warning dialog, choose **Replace** to replace the connection to the previous source node).

*Figure 5. Adding the second month's data*



1. On the Model tab of the Bayesian Network node, for Model name, select **Custom** and enter Jan-Feb in the text box.
2. Select **Continue training existing model**.
3. Click **Run**. The model nugget overwrites the existing one in the stream, but is also added to the Models palette in the upper-right corner.

*Figure 6. Retraining the model*



**Evaluating the Model**

To compare the models, you must combine the two datasets.

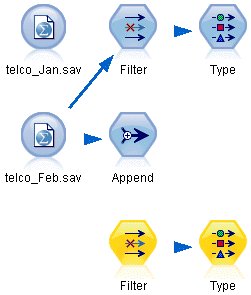
1. Add an Append node and attach both the *telco\_Jan.sav* and *telco\_Feb.sav* source nodes to it.

*Figure 1. Append the two data sources*



1. Copy the Filter and Type nodes from earlier in the stream and paste them onto the stream canvas.
2. Attach the Append node to the newly copied Filter node.

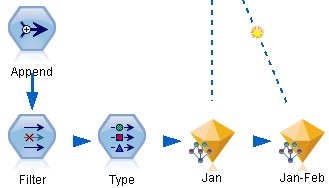
*Figure 2. Pasting the copied nodes into the stream*



The nuggets for the two Bayesian Network models are located in the Models palette in the upper-right corner.

1. Double-click the Jan model nugget to bring it into the stream, and attach it to the newly copied Type node.
2. Attach the Jan-Feb model nugget already in the stream to the Jan model nugget.
3. Open the Jan model nugget.

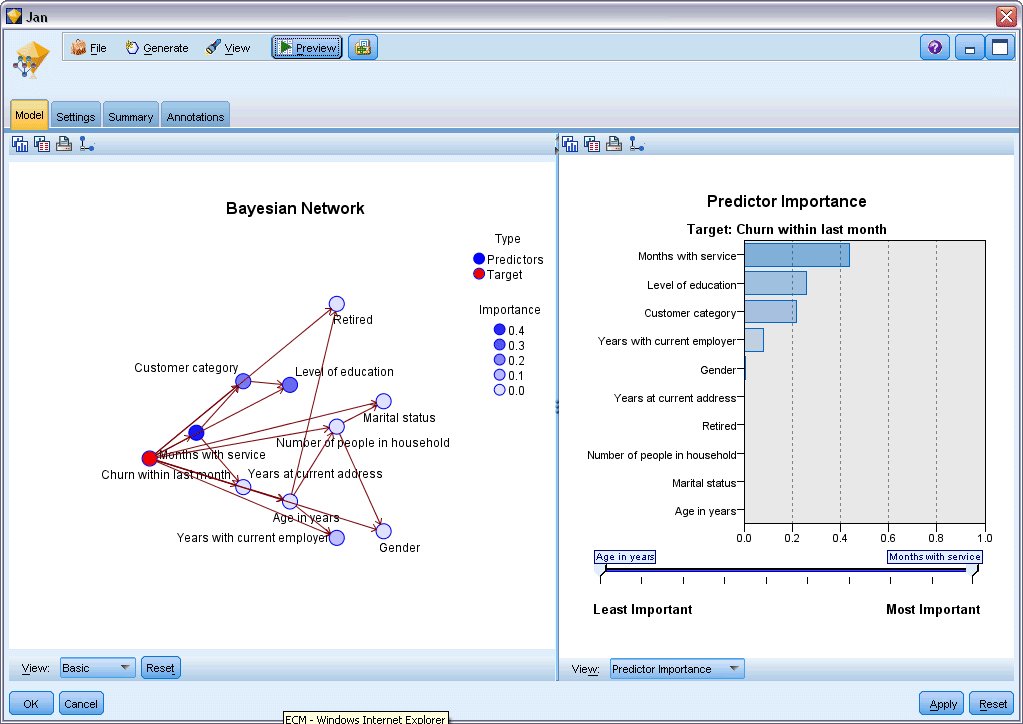
*Figure 3. Adding the nuggets to the stream*



The Bayesian Network model nugget Model tab is split into two columns. The left column 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 column 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.

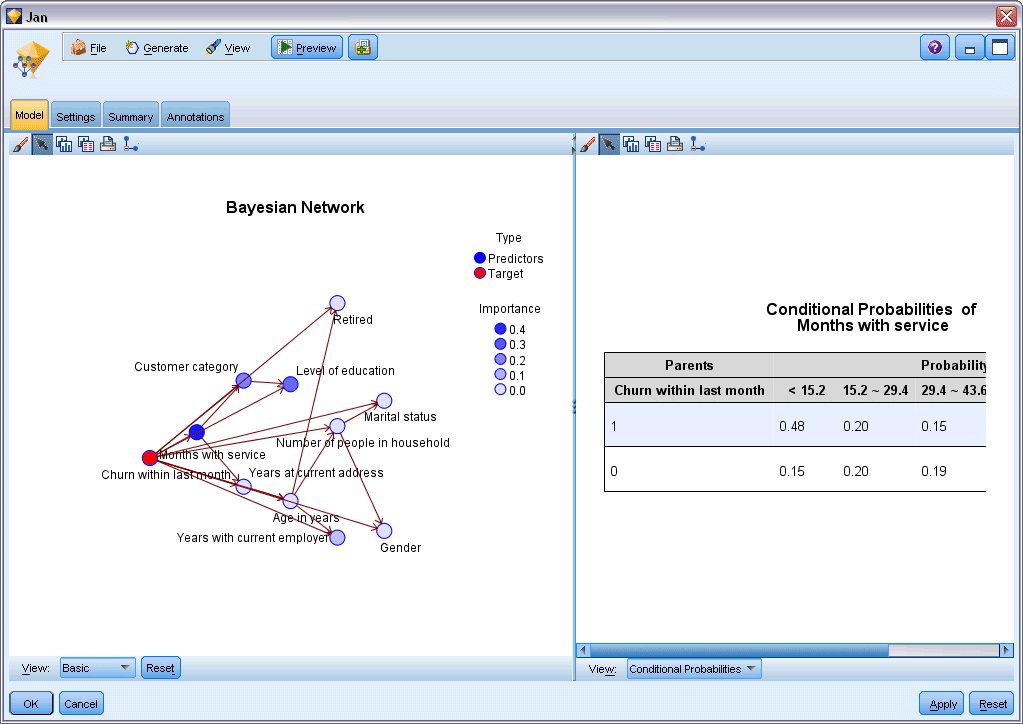
*Figure 4. Bayesian Network model showing predictor importance*



To display the conditional probabilities for any node, click on the node in the left column. The right column is updated to show the required details.

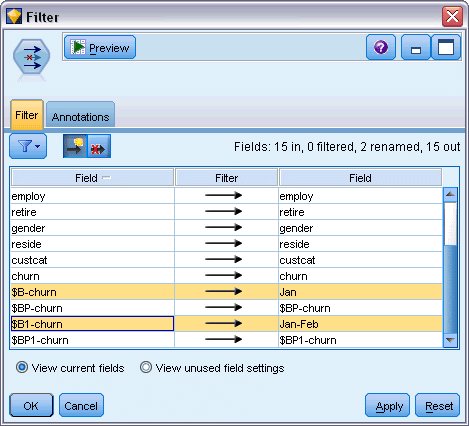
The conditional probabilities are shown for each bin that the data values have been divided into relative to the node's parent and sibling nodes.

*Figure 5. Bayesian Network model showing conditional probabilities*



1. To rename the model outputs for clarity, attach a Filter node to the Jan-Feb model nugget.
2. In the right *Field* column, rename $B-churn as Jan and $B1-churn as Jan-Feb.

*Figure 6. Rename model field names*

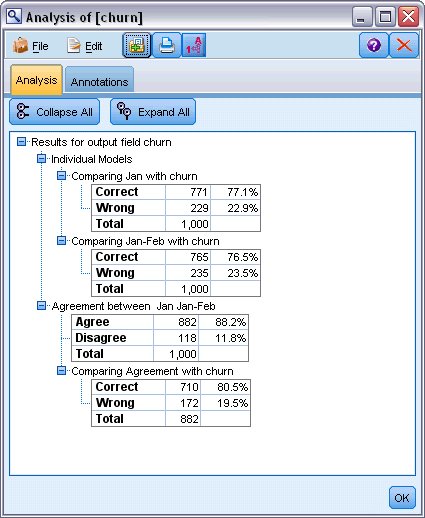


To check how well each model predicts churn, use an Analysis node; this shows the accuracy in terms of percentage for both correct and incorrect predictions.

1. Attach an Analysis node to the Filter node.
2. Open the Analysis node and click **Run**.

This shows that both models have a similar degree of accuracy when predicting churn.

*Figure 7. Analyzing model accuracy*



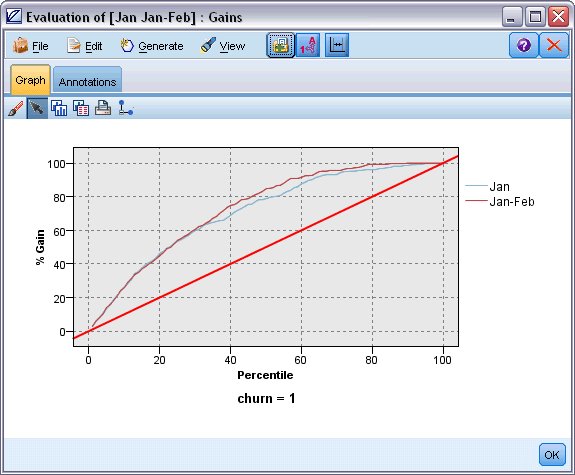
As an alternative to the Analysis node, you can use an Evaluation graph to compare the models' predicted accuracy by building a gains chart.

1. Attach an Evaluation graph node to the Filter node.

and execute the graph node using its default settings.

As with the Analysis node, the graph shows that each model type produces similar results; however, the retrained model using both months' data is slightly better because it has a higher level of confidence in its predictions.

*Figure 8. Evaluating 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.