

Naive Bayes

JMP PRO Use this predictive modeling technique to predict a categorical outcome (classify) as a function of multiple predictor variables. The technique classifies observations by applying Bayes' Theorem to conditional probabilities.

Naive Bayes

1. From an open JMP® table, select **Analyze > Predictive Modeling > Naive Bayes**.
2. Select a nominal or ordinal response variable from **Select Columns** and click **Y, Response**.
3. Select candidate predictor variables and click **X, Factor**.
4. If desired, enter the **Validation Portion** or select a validation column and click **Validation** as was used in this illustration.
5. Click **OK**. JMP displays:

- Total misclassification counts and rates.
- Confusion Matrix detailing the classification performance.
- ROC curves and AUC values (only results for the training data shown here).

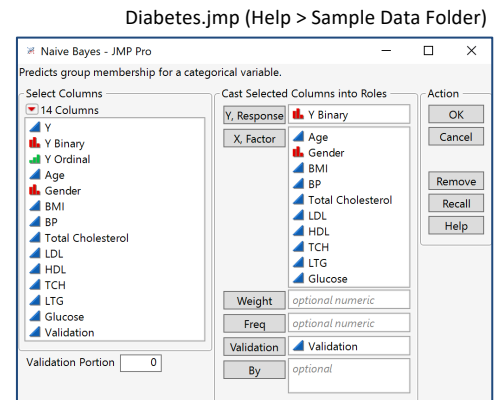
Note: Click on the **top red triangle** and select **Profiler** to display an interactive tool that shows the predicted probability of each class as a function of the levels of the predictor variables. Other options, such as **Assess Variable Importance** can be accessed from the **red triangle** next to Prediction Profiler. The profiler for the three predictor variables HDL, BMI, and LTG shown on bottom right.

Results of the Naive Bayes classifier to predict the level (Low/High) from the 442 diabetes patients:

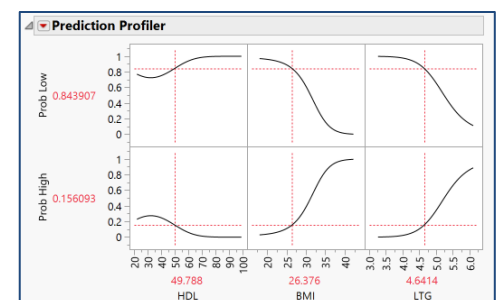
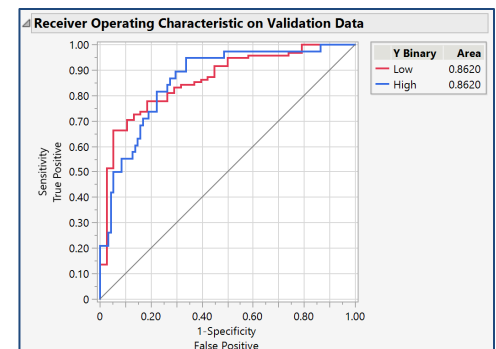
- There are 309 observations in the Training Data. Of these, 66 (21%) were misclassified. $41/(185+41) = 18\%$ of the Low observations were misclassified as High. $25/(25+58) = 30\%$ of the High observations were misclassified as Low.
- There are 133 observations in the Validation Data. Of these, 32 (24%) were misclassified. $25/(70+25) = 26\%$ of the Low observations were misclassified as High. $7/(7+31) = 18\%$ of the High observations were misclassified as Low.

Notes:

Additional options, such as **Lift Curves**, **Save Predicteds**, **Save Prediction Formula**, **Save Probability Formula**, as well as **Publish Probability Formulas** are accessible from the **top red triangle**.



Training					Validation				
		Misclassification	Rate	Misclassifications			Misclassification	Rate	Misclassifications
		309	0.21359	66			133	0.24060	32
Misclassifications									
66									
Confusion Matrix									
Training					Validation				
		Predicted Count					Predicted Count		
Actual Y Binary	Low	High		Actual Y Binary	Low	High	Actual Y Binary	Low	High
Low	185	41		Low	70	25	Low	70	25
High	25	58		High	7	31	High	7	31
		Predicted Rate					Predicted Rate		
Actual Y Binary	Low	High		Actual Y Binary	Low	High	Actual Y Binary	Low	High
Low	0.819	0.181		Low	0.737	0.263	Low	0.737	0.263
High	0.301	0.699		High	0.184	0.816	High	0.184	0.816



Visit **Predictive and Specialized Models > Naive Bayes** in **JMP Help** to learn more.