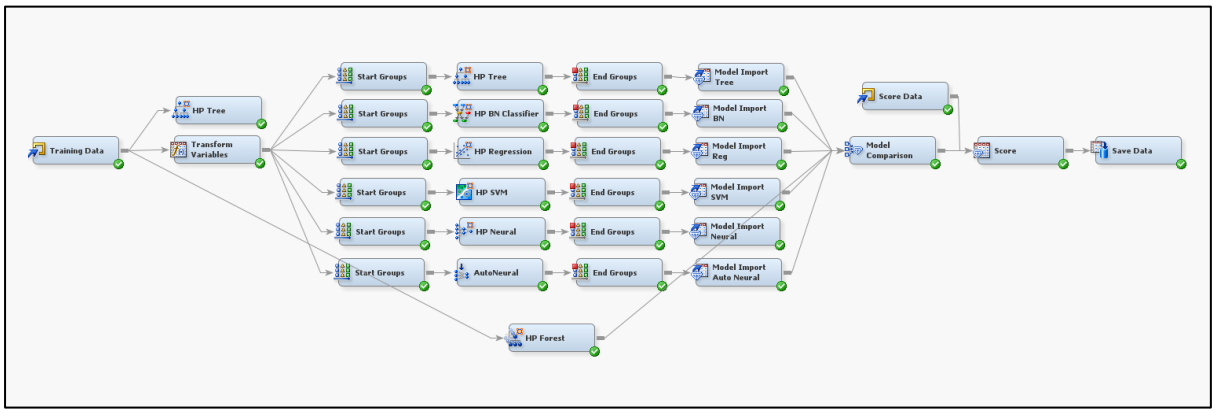
Model building



In SAS Enterprise Miner, the training data was imported in and connected to the "Transform Variables" node as shown in Figure 3.6. The "Transform Variables" node creates k-fold cross validation indicator which divides the data randomly into k-folds. The k value is set to 10.

Each model has their own "Start Groups" and "End Groups" nodes. These were required because the data was segregated into 10 segments of data. The algorithm nodes were placed in between "Start Groups" and "End Groups" node. The algorithms used were Decision Tree, Naïve Bayes, Logistic Regression, SVM, Neural Network, Auto Neural and Random Forest. All the algorithm was high performance node except for Auto Neural. Basically, high performance node offers shorter operation time, which is suitable for larger datasets. The "End Groups" produces the correct score code in order to calculate the cross-validation error. Next, a "Model Import" node is attached to the end to import out the result of the model. All the models generated were compared by using a "Model Comparison" node. The criteria for the best model is based on the lowest misclassification rate. The best model was then tested using score data by using a "Score" node, followed by exporting the scored results for evaluation.