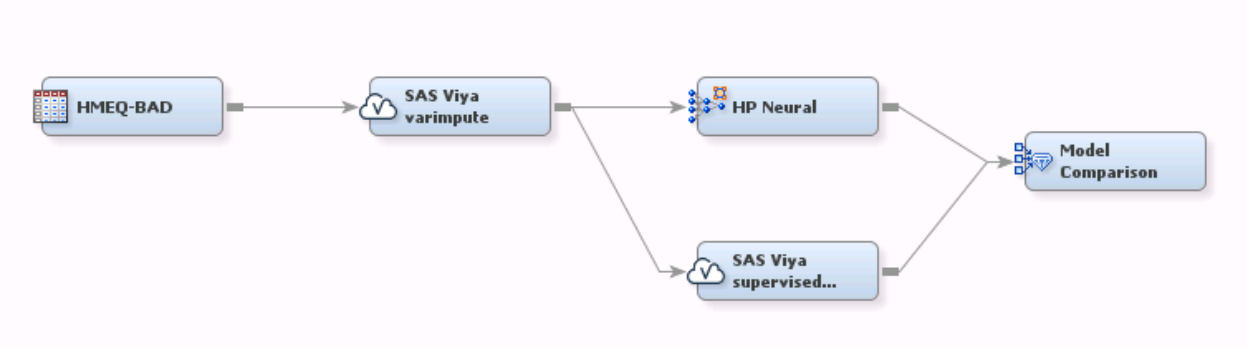
Comparing Models in SAS® Viya™ Using SAS® Enterprise Miner™

Process Flow Diagram:



Data:
The Home Equity (HMEQ) data set in the SAMPSIO SAS library is used to create the data source. You can generate the data in SAS Enterprise Miner 14.3 by selecting **Help > Generate Sample Data Sources** to open the Generate Sample Data Sources window and then selecting **Home Equity**. The target is the binary variable **BAD,** which indicates whether an applicant defaulted (or was seriously delinquent) on a home equity loan, or paid the loan on time.

Goal:
The goal is to compare the best supervised model that is trained by a SAS Viya Code node with a neural network model that is trained by the HP NEURAL procedure, a high-performance SAS Enterprise Miner procedure. Before the models are trained, the missing values for nominal and interval variables are imputed.

Flow:
The Input Data node (titled "HMEQ-BAD" in the flow diagram) represents the data source and provides details about the variables in the data source.

The Input Data node is connected to a SAS Viya Code node (titled "SAS Viya varimpute" in the flow diagram). The SAS Viya Code node enables you to write your own code. The SAS Viya varimpute node performs imputation via the VARIMPUTE procedure (a statistical procedure in SAS® Visual Data Mining and Machine Learning). INPUT statements in PROC VARIMPUTE request that mean imputation be used for the interval variables and median imputation be used for nominal variables. After you run the SAS Viya Code node, the **Variables Imputation Information** table in the **Results** shows the types of imputation methods and the imputed values for each variable.

The HP Neural node trains a multilayer neural network. The **Input Standardization** property, which specifies the method for standardizing interval input variables, is set to **Range**. This property specifies the method that is used to standardize interval input variables. The **Architecture** property, which specifies the type of the network architecture, is set to **One Layer**. Both the number of hidden neurons and the number of hidden layers are set to 3. Other possible values are integers between 1 and 10, inclusive.

The second SAS Viya Code node (titled “SAS Viya supervised learning” in the flow diagram) runs several machine learning models that use the following procedures in SAS Visual Data Mining and Machine Learning: the NNET procedure to train a neural network model, the SVMACHINE procedure to train a support vector machine model, and two TREESPLIT procedures to fit regression trees. The %EM_VIYA_ASSESS macro is used to assess the models. The NAME= argument in this macro is used to specify a name for each model. Because you are assessing multiple models, you need to use the EM_VIYA_MODELSELECTION macro to compare the models and select the best model based on the criteria specified in the SAS Viya Code node.

Because the target variable is binary, the **Class Selection Statistic** property is used to obtain the fit statistic to be used to evaluate the models. The **Model Selection** table in the **Results** shows that the SVM model was selected based on the **Kolmogorov-Smirnov statistic** in the **Train** partition. The SVM model is exported to the Model Comparison node in the flow.

The Model Comparison node compares the performance of the two competing models: the neural network model (from the HP Neural node) and the SVM model (from the SAS Viya Code node). The **Selection Statistic** property is set to **Default**. The champion model (neural network) is selected based on the **Misclassification Rate** criterion in the **Train** partition.