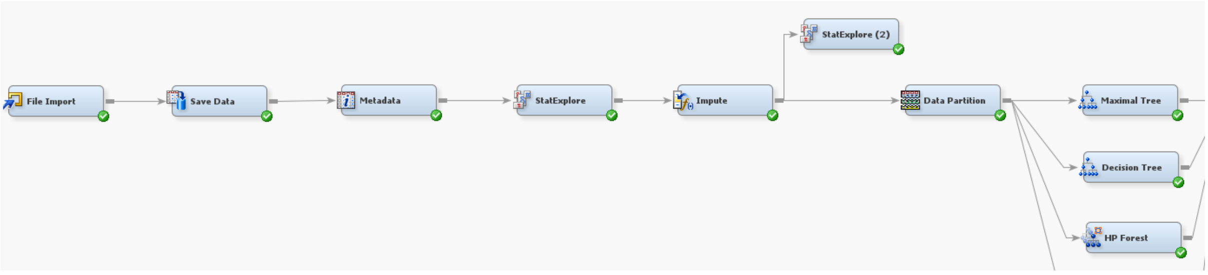
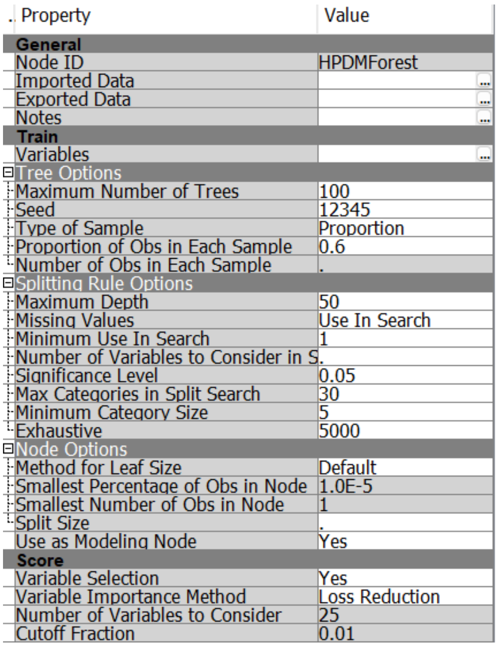
6.0 Ensemble Methods using SAS Enterprise Miner

6.1 Bagging

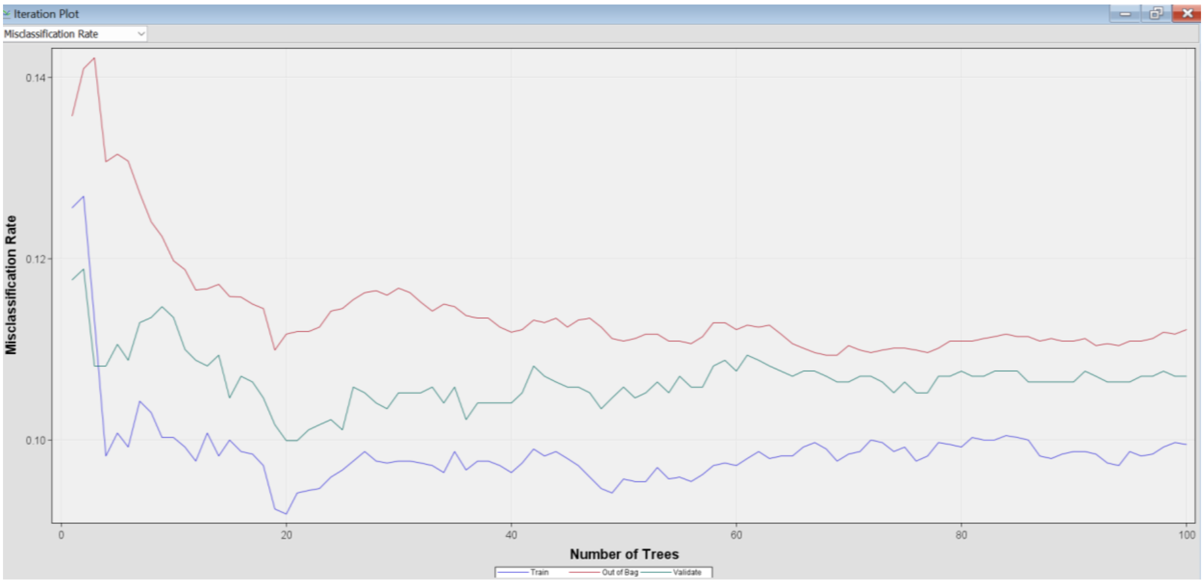
Create a model for Bagging using “HP Forest” node. Keep default settings.



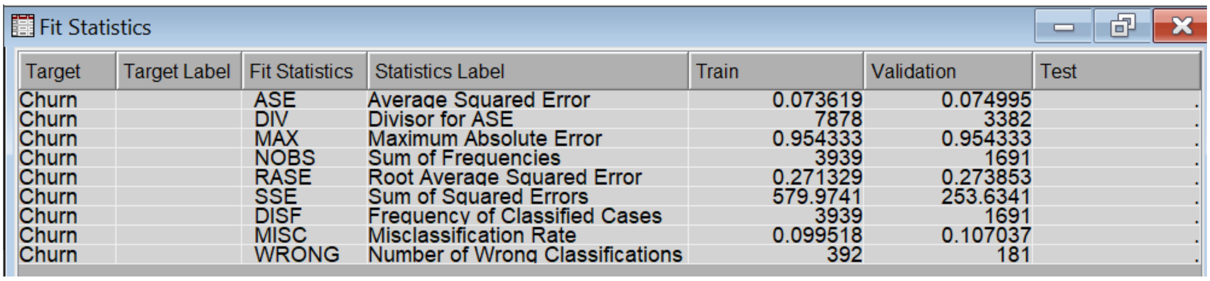


Property	Value
General	
Node ID	HPDMForest
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
⊟Tree Options	
Maximum Number of Trees	100
Seed	12345
Type of Sample	Proportion
Proportion of Obs in Each Sample	0.6
Number of Obs in Each Sample	.
⊟Splitting Rule Options	
Maximum Depth	50
Missing Values	Use In Search
Minimum Use In Search	1
Number of Variables to Consider in S	.
Significance Level	0.05
Max Categories in Split Search	30
Minimum Category Size	5
Exhaustive	5000
⊟Node Options	
Method for Leaf Size	Default
Smallest Percentage of Obs in Node	1.0E-5
Smallest Number of Obs in Node	1
Split Size	.
Use as Modeling Node	Yes
Score	
Variable Selection	Yes
Variable Importance Method	Loss Reduction
Number of Variables to Consider	25
Cutoff Fraction	0.01

Based on Iteration Plot, misclassification rate plateaued out when number of trees reaches 20.



Based on Fit Statistics, misclassification rate is 0.09952 for training dataset and 0.1070 for validation dataset.

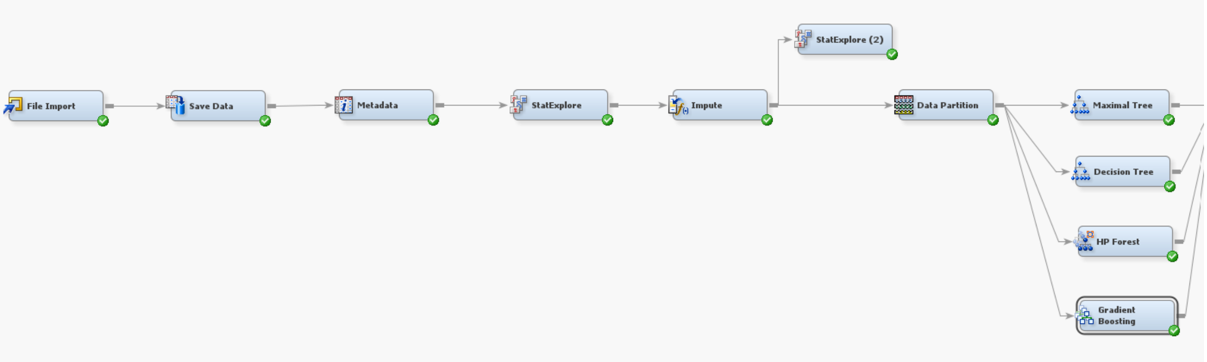


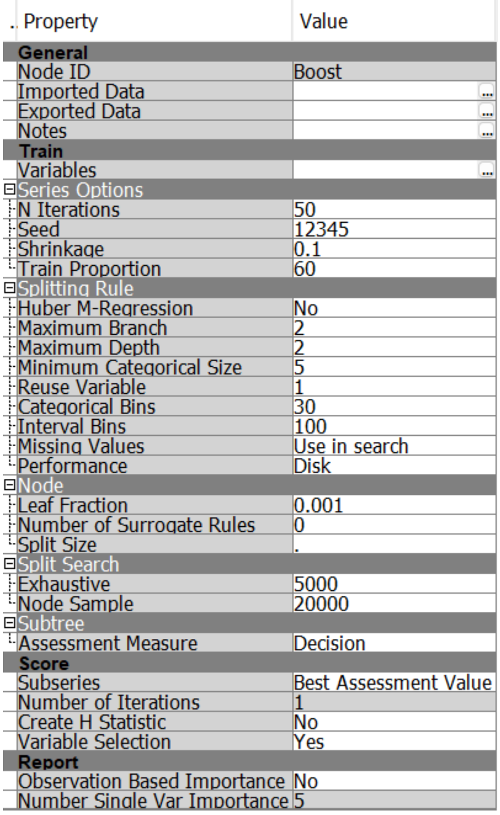
Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		ASE	Average Squared Error	0.073619	0.074995	.
Churn		DIV	Divisor for ASE	7878	3382	.
Churn		MAX	Maximum Absolute Error	0.954333	0.954333	.
Churn		NOBS	Sum of Frequencies	3939	1691	.
Churn		RASE	Root Average Squared Error	0.271329	0.273853	.
Churn		SSE	Sum of Squared Errors	579.9741	253.6341	.
Churn		DISF	Frequency of Classified Cases	3939	1691	.
Churn		MISC	Misclassification Rate	0.099518	0.107037	.
Churn		WRONG	Number of Wrong Classifications	392	181	.

6.2 Boosting

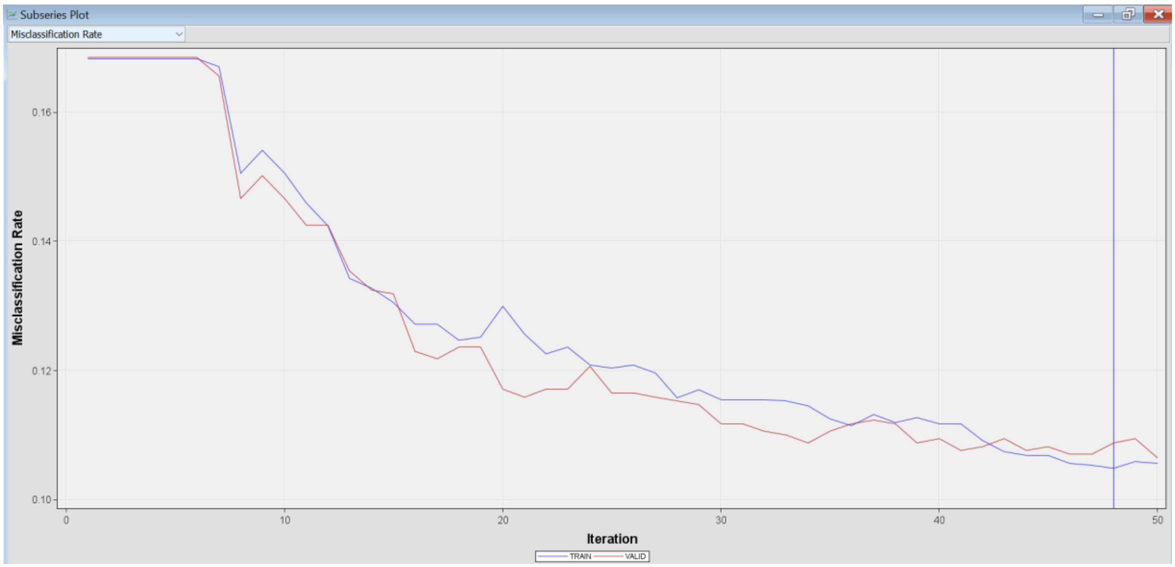
Create a model for Boosting using “Gradient Boosting” node. Keep default settings.



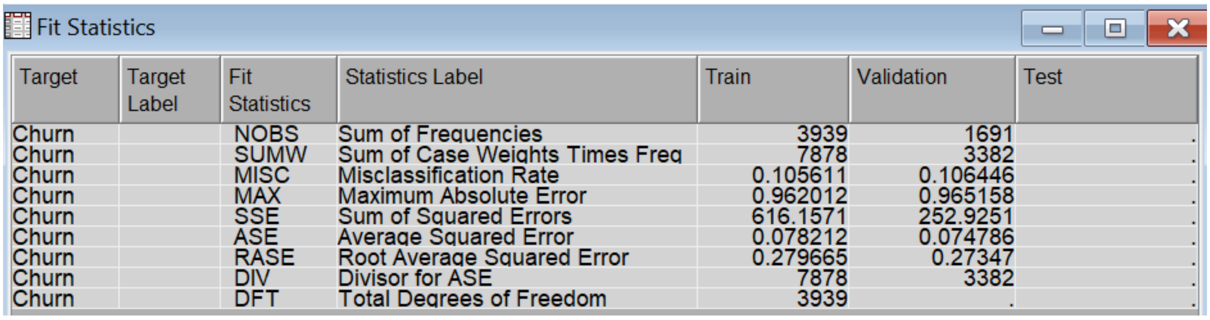


Property	Value
General	
Node ID	Boost
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
⊟Series Options	
N Iterations	50
Seed	12345
Shrinkage	0.1
Train Proportion	60
⊟Splitting Rule	
Huber M-Regression	No
Maximum Branch	2
Maximum Depth	2
Minimum Categorical Size	5
Reuse Variable	1
Categorical Bins	30
Interval Bins	100
Missing Values	Use in search
Performance	Disk
⊟Node	
Leaf Fraction	0.001
Number of Surrogate Rules	0
Split Size	.
⊟Split Search	
Exhaustive	5000
Node Sample	20000
⊟Subtree	
Assessment Measure	Decision
Score	
Subseries	Best Assessment Value
Number of Iterations	1
Create H Statistic	No
Variable Selection	Yes
Report	
Observation Based Importance	No
Number Single Var Importance	5

Based on Iteration Plot, misclassification rate plateaued out at 48th iteration.



Based on Fit Statistics, misclassification rate is 0.1056 for training dataset and 0.1064 for validation dataset.


Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
Churn		NOBS	Sum of Frequencies	3939	1691	.
Churn		SUMW	Sum of Case Weights Times Freq	7878	3382	.
Churn		MISC	Misclassification Rate	0.105611	0.106446	.
Churn		MAX	Maximum Absolute Error	0.962012	0.965158	.
Churn		SSE	Sum of Squared Errors	616.1571	252.9251	.
Churn		ASE	Average Squared Error	0.078212	0.074786	.
Churn		RASE	Root Average Squared Error	0.279665	0.27347	.
Churn		DIV	Divisor for ASE	7878	3382	.
Churn		DFT	Total Degrees of Freedom	3939	.	.

6.3 Model Comparison

Compare model performance using “Model Comparison” node.

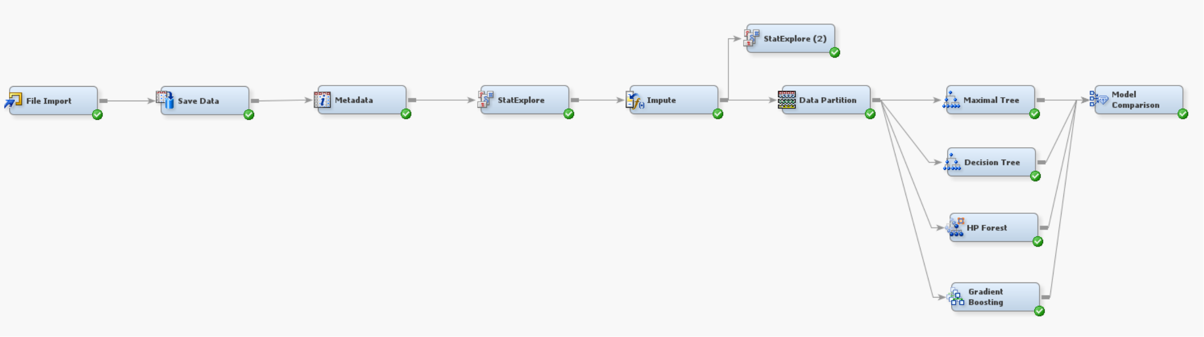
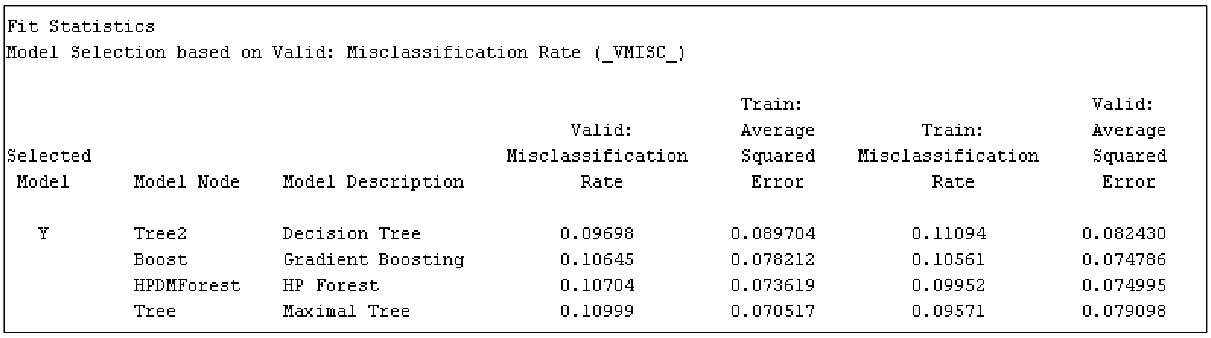


Figure below shows the Fit Statistics for model comparison.



```
Fit Statistics
Model Selection based on Valid: Misclassification Rate (_VMISC_)

                                                                Train:                            Valid:
                                                  Valid:        Average         Train:            Average
Selected                                    Misclassification   Squared   Misclassification       Squared
 Model      Model Node    Model Description        Rate          Error           Rate              Error

   Y        Tree2         Decision Tree          0.09698       0.089704        0.11094           0.082430
            Boost         Gradient Boosting      0.10645       0.078212        0.10561           0.074786
            HPDMForest    HP Forest              0.10704       0.073619        0.09952           0.074995
            Tree          Maximal Tree           0.10999       0.070517        0.09571           0.079098
```

The selected model is pruned Decision Tree (Tree2 in the figure) with a validation misclassification rate of 0.09698 or 9.698%. Although the two ensemble methods helped reduce the training misclassification rate, ensemble methods resulted in higher validation misclassification rate than the pruned Decision Tree. This outcome is not uncommon and can be due to various reasons:

1. **Overfitting in Ensemble Methods:**
   - While ensemble methods (such as Random Forest or Gradient Boosting) aim to reduce overfitting, improper tuning or inadequate control over model complexity might lead to overfitting the training data. This can result in poorer performance on unseen validation data.
2. **Sensitivity to Hyperparameters:**
   - Ensemble methods often have multiple hyperparameters to tune (e.g., number of trees in Random Forest, learning rate in Gradient Boosting). Suboptimal hyperparameters can negatively affect model performance on validation data.

Nevertheless, the resulting difference is not significant (1 - 2%). It is fair to conclude that all the models including pruned Decision Tree, Random Forest and Gradient Boosting managed to deliver good classification accuracy, with small misclassification rate of 9.6 – 10.7%.