**Bankruptcy Prediction**

SUMMARY

INTRO

The Bankruptcy Data describes whether a bankruptcy was filed or not for 5436 observations of 13 variables. The binary outcome (1 or 0) is predicted using the 12 independent variables.

APPROACH

To predict bankruptcy or not(1 or 0) different supervised machine learning models like Logistic Regression, Classification Tree, GAM and Neural network is used. The data is divided into train and test. Every model is then trained using in sample data and the hyper-parameters are tuned. A final model is selected comparing all in sample Asymmetric cost of models. The selected final model is then tested to find out of sample Asymmetric cost.

MAJOR FINDINGS

* Logistic Regression performed the worst with an in sample Asymmetric cost of 0.52
* Neural Network gave the best in sample Asymmetric cost of 0.42
* Logistic Regression, Classification tree, GAM and Neural Networks is the order of in sample performance from worst to best.
* The out of sample Asymmetric cost of Neural Network is 0.43 close to the in sample Asymmetric cost. It is a stable model without overfitting as the performance is consistent in both the test and train data.

### 1.Logistic Regression

* ROC curve of Logistic Regression

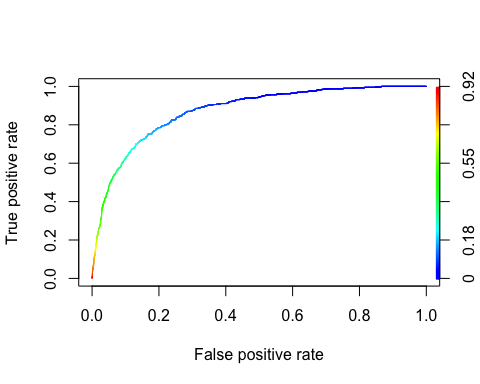


FIG.2.1.1

* AUC of Logistic Regression

## [1] 0.8717709

## Predicted  
## Observed FALSE TRUE  
## 0 1389 2105  
## 1 21 562

* In-Sample metrics of logistics:

## Logistic  
## AUC 0.8717709  
## Asymmetric Rate 0.5214619

### 2.Classification Tree

* ROC curve of Classification Tree

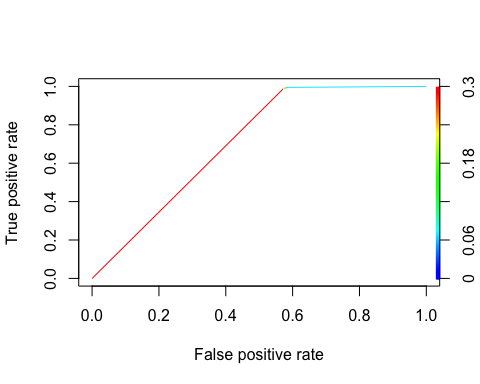


FIG.2.2.1

* AUC of Classification Tree

## [1] 0.7093952

## Predicted  
## Observed FALSE TRUE  
## 0 1446 2048  
## 1 3 580

* In-Sample metrics of Classification Tree:

## Classification.tree  
## AUC 0.7093952  
## Asymmetric Rate 0.5030660

### 3.GAM

* Obtained Smoothing functions of the predictors:

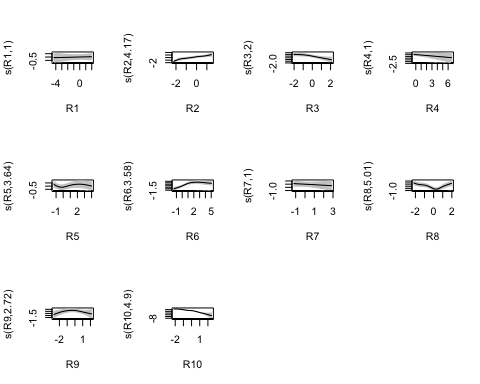


FIG.2.3.1

## Predicted  
## Observed FALSE TRUE  
## 0 1593 1901  
## 1 9 574

* In-Sample metrics of GAM:

## Classification.tree  
## Asymmetric Rate 0.503066

### 

### 4.Neural Net

## # weights: 13  
## initial value 4009.779330   
## iter 10 value 1367.831523  
## iter 20 value 1228.079777  
## iter 30 value 1162.902254  
## iter 40 value 1152.928398  
## iter 50 value 1149.903026  
## final value 1149.901138   
## converged

## Predicted  
## Observed FALSE TRUE  
## 0 1736 1758  
## 1 30 553

* In-Sample metrics of Neural Network:

## Classification.tree  
## Asymmetric Rate 0.4385578

### 

### In sample model comparisons

## Model Asymmetric.Cost  
## 1 Logistic Regression 0.5214619  
## 2 Classification Tree 0.5030660  
## 3 GAM 0.4684817  
## 4 Neural Network 0.4385578

* Final model we have chosen is Neural Network based on Asymmetric Cost

## Data Assymetric.cost  
## 1 In\_Sample 0.4238411  
## 2 Out\_of\_Sample 0.4385578

* The out of sample cost is close to in-sample cost, it can be considered as a stable model without overfitting