Assignment 4 (Project)

05-December-2016

Submitted by : [Group DA-04]

P.N. Vamshi [13EC10044]

Nikhil Kashyap [13EE10033]

Manognya Deepthi G [13IM10008]

Mukesh Sahani [13EC10039]

Manoj Meena [13EE10028]

Customer Churn Prediction

Approach Used

1) Converted attributed having factor as a parameter to numeric values.
2) Handled missing values by **na.omit()** on data frame.
3) Identified and removed unnecessary attributes (eg: Id, state, area_code, phone_number).
4) Calculated correlation matrix by using **cor()** method.
5) Visualize correlation matrix using **heatmap()** function available in R library .
6) Constructed Naive Bayes model, SVM model and C5.0 decision tree model by using **naiveBayes()**, **svm()** and **C5.0()** functions respectively on training data.
7) Splitting the data into training(70%) and test set(30%) by maintaining true to false proportion is done by making use of **sample()** function available in R package library.

8) Predicted the values for the test data using **predict()** function available in R package library.

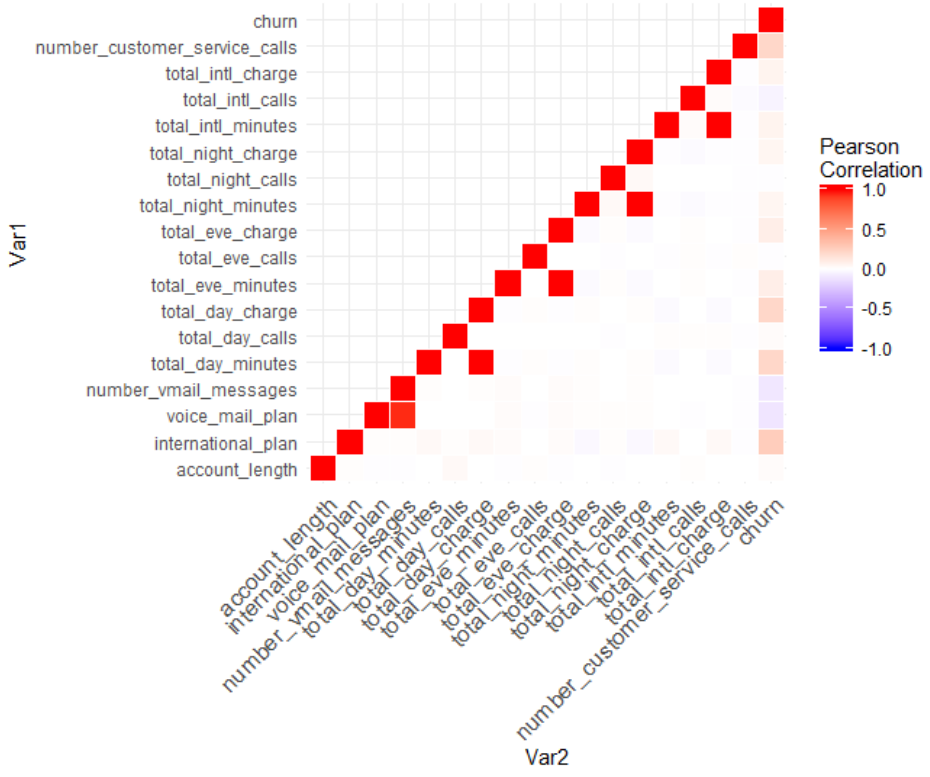
9) Constructed confusion matrix by using predicted values and actual values of the test data of churn attribute by the help of **confusion.matrix()** function.

10) Calculated accuracy, precision and recall values for all the models.

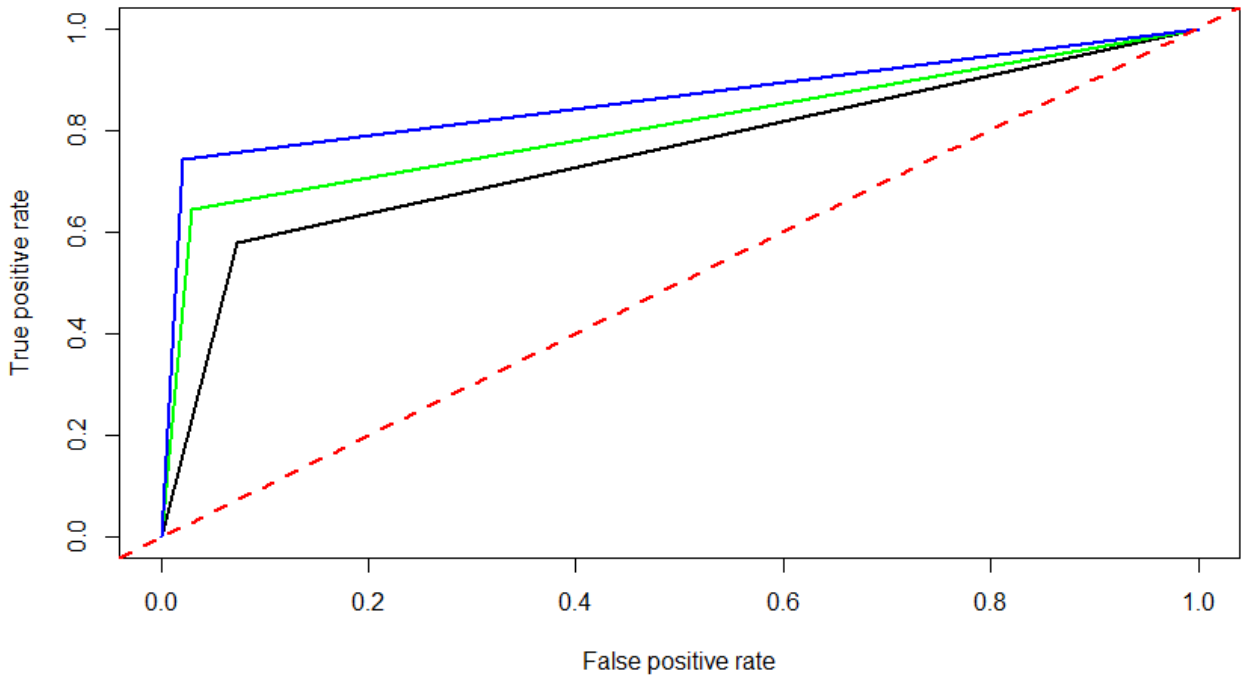
11) Plotted ROC curve for all the models and found that the classifier that is more close to best classifier is C5.0.

12) Performed feature scaling and increased the boosting iteration i.e. trials of C5.0 to increase the accuracy.

Correlation Matrix Plot



ROC Plot



Blue - C5.0 Decision Tree | Green - SVM Model | Black - Naive Bayes Model

Libraries

Libraries that are needed to execute the code are :
library(caret), library(rpart), library(C50), library(rattle), library(party), library(partykit), library(ROCR), library(ggplot2), library(reshape2), library(car), library(corrplot), library(e1071), library(SDMTools)