Evaluation Report

Capstone Term II – AIDI 2005

Facilitator:

Marcos Bittencourt

Date:

Jan 24 2020

Student:

Anmol Panna (100722039)

Sharang Ahluwalia (100700203)

Abdul Rahman Khan (100719011)

Levi Nair (100733132)

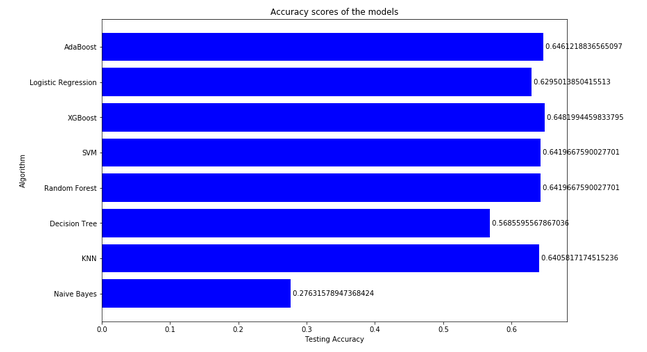
# Key Findings

Various algorithms were fitted into the model, Following are the names of algorithms:

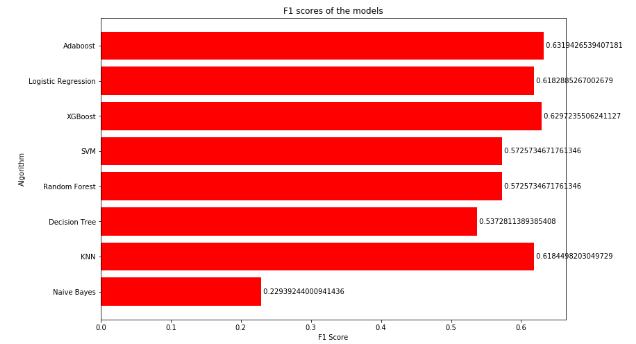
* Naïve Bayes:
  + Fast and easy to predict. Performs well with categorical inputs. As it makes probabilistic predictions, it is used to find the odds.
  + Accuracy and F1 scores were not good
* K- nearest Neighbor:
  + Good for Noisy Data
  + Had to deal with outliers as KNN suffers if there are outliers.(also had to determine value of K)
* Random Forrest:
  + Efficient estimates of test errors without incurring cost of repeated model
  + High computational costs and predictions are slower
* SVM
  + Good with outliers
  + Selecting kernel function was tricky
* XG Boost
  + Extremely fast (parallel computation)
  + But only worked with numeric features and hyperparameter tuning is important otherwise it will cause overfitting
* Logistic Regression
  + Resistant to overfitting
  + Cannot predict continuous outcomes
* ADABoost
  + Sensitive to Noise and outliers

# Scorecards

Accuracy Scores of Algorithm:



F1 Scores of Algorithm:



Best algorithm as per accuracy, F1 score and computational effieciency is:

AdaBoost:

Adaboost was selected because it satisfied our low computation time and high accuracy relative to the other algorithms. Although Adaboost suffers from overfitting, there are many methods to deal with overfitting