Safe Driving Challenge

M Susmitha-17wh1a0591
M Naga Pravallika-18wh5a0517
T Bindhu Bhargavi-18wh5a0520

Department of Computer Science and Engineering
Bvrit Hyderabad College of Engineering for women

June, 2020

TOPICS

INTRODUCTION

- The Safe Challenge project objective is to design a classifier that will detect whether the driver is alert or not alert, employing data that are acquired while driving.
- Safe driving is one important factor for safe living. The objective of this challenge is to design a model that will detect drivers attentiveness from the information acquired while driving.

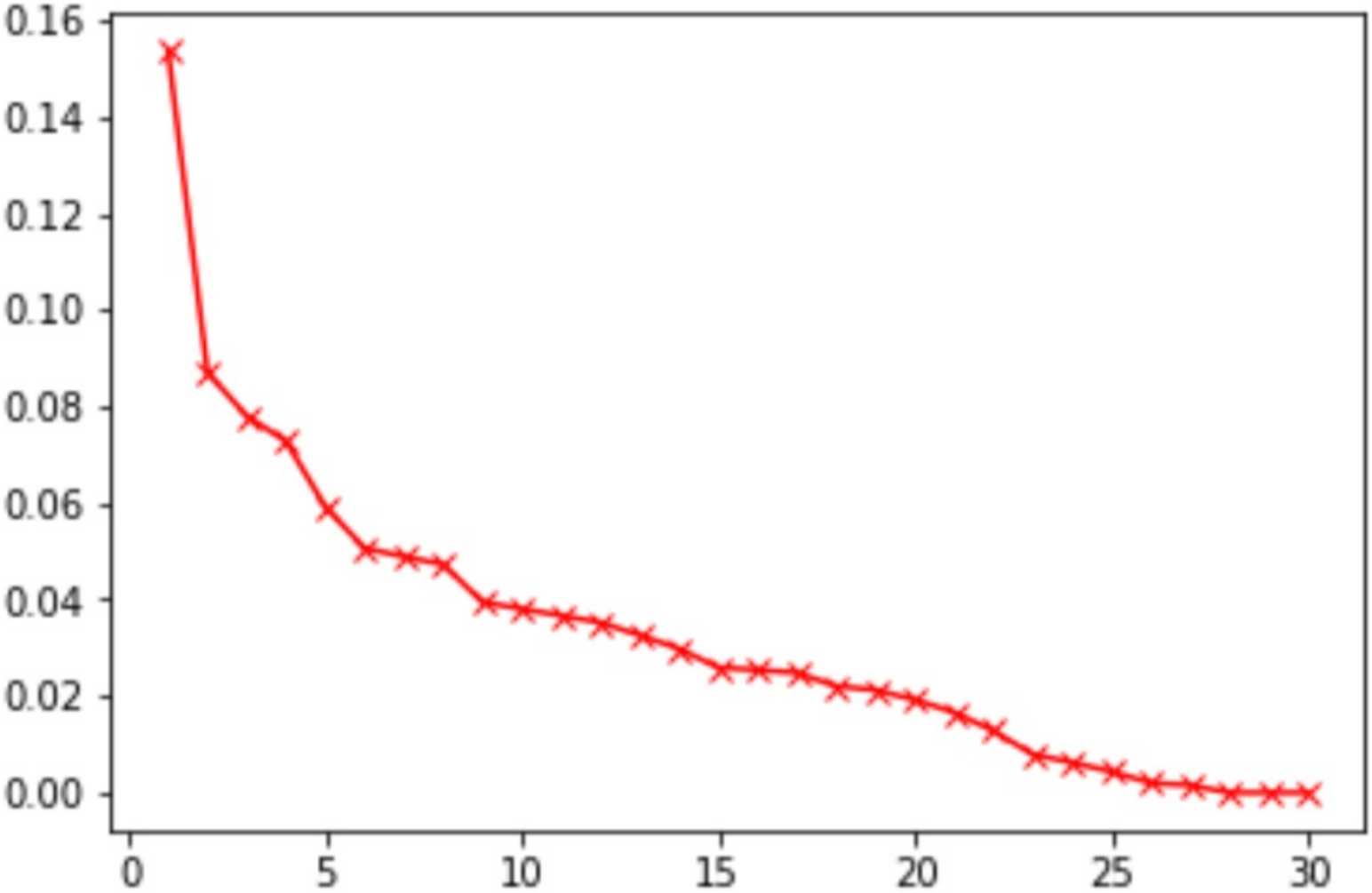
DATA SET

- There are 604,329 instances of data in the training data-set and 120,840 instances of data in the test data-set. The data for this challenge shows the results of a number of trials, each one representing about 2 minutes of sequential data that are recorded every 100 ms during a driving session on the road or in a driving simulator.

DATA PRE-PROCESSING METHOD (MODEL-1)

- There are 30 features included in the data-set, thus filter those features with higher impact could not only save computational resources but also potentially improve the performance of the predictive model. Principle Component Analysis (PCA) is applied as the feature engineering technique in this case.

- x axis - 30 Features
  y axis - Fraction of Variances



ACCURACY

Model	Accuracy	Accuracy Score
Logistic Regression	64.60%	0.597
Nave Bayes	61.74%	0.573
Random Forest	93.54%	0.93
Neural Network	65.01%	0.6023
SVC	64.63%	0.59
NU-SVC	78.63%	0.719
C_SVC	67%	0.625

DATA PRE-PROCESSING METHOD (MODEL-2)

- For these reasons, We decided to use rolling means and standard deviations of each features as new features in- stead of simply using stable means and standard deviations in order to make full use of the sequential feature.
- The rolling window is set to 5, as for every 5 instances (500ms), it calculates the mean and standard deviation for them, then the algorithm drop the first instance and add a new instance, etc.

Model	Accuracy	Accuracy Score
Logistic Regression	61.21%	0.58
Nave Bayes	62.86%	0.66
Random Forest	98.91%	0.98
Neural Network	80.76%	0.77

MODEL OPTIMIZATION AND SELECTION

- Though the neural network model fails to produce a better performance than the random forest model, it is still not convincing that random forest is always the best option for this problem. Neural network still shows great potential to produce good result. Further work could to optimize the architecture of neural networks.

SUMMARY

- This project was meant to build a supervised learning model to predict not alert drivers, the model with the best performance is achieved by Random Forest with 50 trees in it. It predicts 16671 of 29914 not alert drivers correctly in the test data. It reaches a classification accuracy of 81.86

SOURCES I

 Kaggle
https://www.kaggle.com/c/stayalert/overview