

Forecast of Vehicles Accident Severity

Audience

The project is driving our examination is the comprehension of explicit conditions that influence the seriousness of a vehicle crash. The reason for this undertaking is to feature effective factors while driving a vehicle to protect from accidents.

Objective

This project is showing different kind of accident in UK where collects 3 years dataset 2014, 2015, and 2015. The methodologies will cover in this project decision tree, Neural n/w, Logical Regression. It will help predicting some causes of accident where taking steps might prevent accidents.

Dataset

The data sets are built in the following aspects:

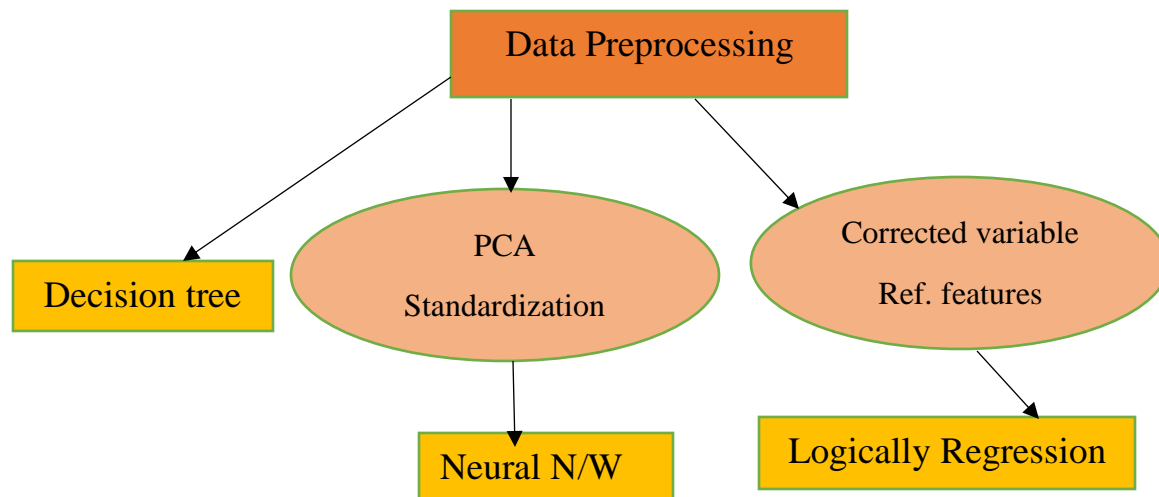
- Severity
- Sex of casualty
- Age of casualty
- Casualty class
- Type of vehicle
- Number of vehicles
- Weather condition
- Lighting condition
- Time 24 hours
- Road surface
- Experiment
- Reference Number
- Grid Ref: Easting
- Grid Ref: Northing

Dataset source:

<https://data.gov.uk/dataset/6efe5505-941f-45bf-b576-4c1e09b579a1/road-traffic-accidents>

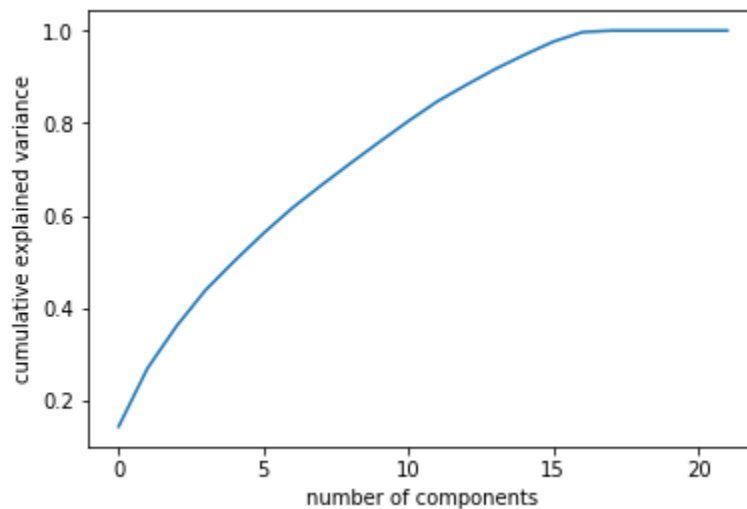
The methodology as follows:

A. Data preprocessing



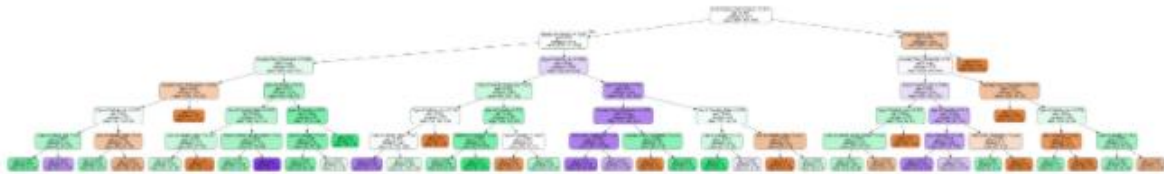
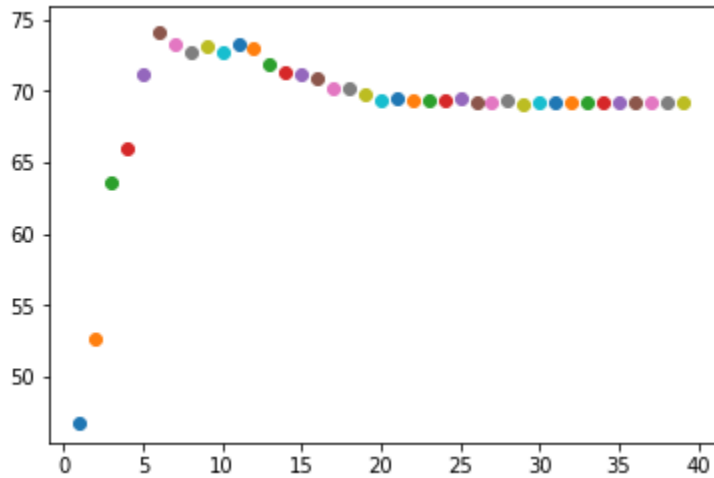
- The three datasets were merged
- Deleting the missing value data
- Deleting some columns which was not necessary

B. PCA



It contains first 12 components which covers mostly variance

C. Decision tree



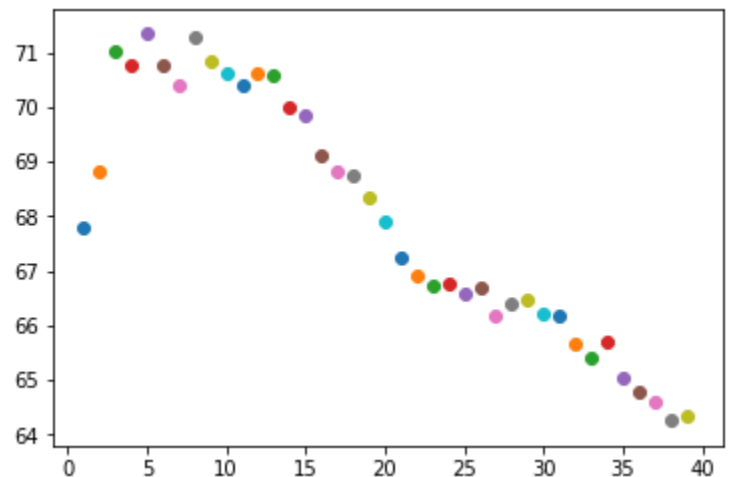
On the above the graph depicts that the best accuracy based on the customized datasets and the tree also shows the importance of road surface, casualty of sex etc.

Selected forest: The standard deviation is 3.98%

D. Neural N/W

- It contains 24 nodes
- Maintaining Standardized data via PCA
- The mean is 72.66%

E. K-Nearest Neighbors



- It also uses standardized via PCA
- On the above the graph predicts the number of neighbors which is the best accuracy based on the customized datasets
- The k-fold means= 71.37%

F. Logically the Regression

- According to PCA the mean is 53.12%

G. Conclusion

Algorithm	Mean Accuracy	Standard Deviation
Decision Tree	73.95	2.64
Random Forest	67.15	3.98
Neural Network	72.66	2.95
KNN	71.37	2.77
Logistic Regression with PCA	53.12	2.32
Logistic Regression without PCA	54.09	3.03

On the above, the final model decision tree where mean is 73.95%, therefore we can say that the most 3 features involve the severity of vehicle accident are – Road Surface dry, road surface wet, and Casualty of Sex.