

Predicting the severity of Car Accident

Kushagra Kumar Rajput

September 5, 2020

1. Introduction

1.1 Background

Today owning a car is no longer a luxury in-fact it is now a necessity of having a car in every household. As the number of cars is increasing day by day on the road the risk of accidents is also increasing. Accidents not only cause human life loss, but financial losses are also incurred. All accidents do not have the same intensity or severity and losses occur vary largely based on the severity of the accident. The severity helps the government and other agencies to predict the number of serious patients that are going to be there. Moreover, with the help of severity government can also plan to work in those areas where frequently more severe accidents are going to occur based on the location of accidents. Furthermore, severity prediction can also help in planning for the weather conditions when accidents are more prone to be severe.

1.2 Problem

Data that might contribute in determining the severity of the accident might include the number of people involved in the event, number or type vehicles involved in the accident, how were the weather conditions at the time of the incident and many more. This project aims to predict the severity of the accident based on these data.

1.3 Interest

The Government and the insurance companies would be very much interested in the prediction of severity of the accidents as it will help in better planning the road safety so that number of severe accidents can be reduced.

2. Data acquisition and cleaning

2.1 Data Sources

There is a large amount of data present on the internet regarding road accidents, but to find the data set to meet the project requirement which to find data set with severity was a bit difficult. Finally, after a lot of google search, I decided to move forward with data on Kaggle.com provided by the Department of Transport UK.

2.2 Data Cleaning

Data used in this project was downloaded from the Kaggle.com which was published by the Department of Transport UK. There were some of the rows which were missing the values, so I

decided to remove those entries from the data as I already a large amount of data even after removing those entries.

Since no data set is perfect as it is and the same is true for this data set a well. Many attributes would not help to learn and hence predict the severity of the accident.

2.3 Feature Selection

After cleaning the data, removing all the null values and the redundant data there were 607216 data samples and 31 features in the data set. The definition of each feature was looked upon and the best features were chosen from the feature set that is going to help in determining the target variable i.e. accident severity.

The list of dropped features is:

- Accident_Index
- Location_Easting_OSGR
- Location_Northing_OSGR
- Longitude
- Latitude
- Date
- Time
- Local_Authority_.Highway
- Local_Authority_.District
- X1st_Road_Number
- X2nd_Road_Number
- Did_Police_Officer_Attend_Scene_of_Accident
- LSOA_of_Accident_Location

All these features were not considered in calculating the severity of the accidents as either these were some unique id given to the data or date, time etc.

Following is the list of features that were used to train the model:

- Police_Force
- Number_of_Vehicles
- Number_of_Casualties
- Day_of_Week
- X1st_Road_Class
- Road_Type
- Speed_limit
- Junction_Detail
- Junction_Control
- X2nd_Road_Class
- Pedestrian_Crossing.Human_Control
- Pedestrian_Crossing.Physical_Facilities

- Light_Conditions
- Weather_Conditions
- Road_Surface_Conditions
- Special_Conditions_at_Site
- Carriageway_Hazards
- Urban_or_Rural_Area