Car Accident Severity

## ****Data****

A dataset of UK traffic data from 2005 and 2015 from Kaggle was used. It is amassed by the UK Government recording over 1.7 million accidents.

The original dataset on Kaggle has the following columns:

Accident\_Index

Location\_Easting\_OSGR

Location\_Northing\_OSGR

Longitude

Latitude

Police\_Force

Accident\_Severity

Number\_of\_Vehicles

Number\_of\_Casualties

Date

Day\_of\_Week

Time

Local\_Authority\_(District)

Local\_Authority\_(Highway)

1st\_Road\_Class

1st\_Road\_Number

Road\_Type

Speed\_limit

Junction\_Detail

Junction\_Control

2nd\_Road\_Class

2nd\_Road\_Number

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

Did\_Police\_Officer\_Attend\_Scene\_of\_Accident

LSOA\_of\_Accident\_Location

The following features were dropped due to significant portion of missing data of irrelevant to the analysis.

Location\_Easting\_OSGR

Location\_Northing\_OSGR

LSOA\_of\_Accident\_Location

Carriageway\_Hazards

Did\_Police\_Officer\_Attend\_Scene\_of\_Accident

Special\_Conditions\_at\_Site

The dataset has an initial total of 1,780,653 samples, each with 32 different features. The feature we want to predict is Accident\_Severity and it is a categorical feature that takes the values 1, 2, 3 to indicate that the crash was of class Fatal, Serious, or Slight, respectively. The remaining features are both numerical and categorical. Additionally, there’s also four geo spatial features used to indicate longitude and latitude of the crash site.