# UK Road Accidents
# Risk Factors

**Yash Karle**

# Table of Contents

# List of Figures

# 1. Introduction

1) Problem definition and ideation

   The aim of this capstone project is to predict the severity (slight, serious, fatal) of personal injury accidents on public roads in the UK. Firstly, by training a supervised learning multi-class classification we aim to achieve this objective. Secondly, we plan to identify high risk factors from the labelled dataset which are essentially top predictor variables for the serious and fatal accidents thereby aiding multiple use-cases like preventive measures, alerting systems and validating injury claims as discussed in detail below. Lastly, the bigger picture purpose of the project would be to drive data-driven enhancement of the road and vehicle conditions that if executed correctly has the potential to save hundreds of thousands of lives by taking simple preventive measures.

2) Target audience, stakeholder personification

   The intended target audience are stakeholders ranging from government public road department to the vehicle/health insurance firms where the risk factors pointed out by the model would enable the stakeholders to make data-driven decisions for road development planning and optimize insurance policy rates respectively.

   Some of the sample questions that the stakeholders would be seeking answers for and would help with their decision making:
   - What causes most road accidents in the UK?
   - What car is in the most accidents in the UK?
   - Which color car has the most accidents in the UK?
   - Where do most accidents happen in the UK?
   - What type of roads have the most accidents?
   - Which type of road is statistically the safest?
   - What is the biggest cause of road accidents in the UK?
   - What is the risk of dying in a car accident in the UK?
   - What role does weather conditions play in the accidents?

3) Intended use-cases and benefits

   The output of the project would focus on the following specific use-cases and attempt to hypothesize and/or form simulations for planning and optimizing for the above-mentioned stakeholders
   - What are the key risk factors (road conditions, vehicle characteristics, any other) in road accidents?
   - Can we accurately predict the severity of accidents given values for these factors?
   - Could this be used to validate injury claim applications?
   - What easy preventive actions could help in curbing the risk of more severe road accidents?
   - Could authorities use monitoring and alerting systems to alert travelers when high-risk factors are evident?

## 2. Data

1) Data source used: UK Road Accidents **(data.gov.uk, 2020)**
2) Data frequency, duration: Yearly data from 2015-2019 (5 years)
3) Data entities (indexes): Accidents (acc_id), Vehicle (veh_ref, acc_id), Casualty (cas_ref, acc_id, veh_ref)
4) Data variables: Please find below the variables listed for each of these entities

| Accident Circumstances | Vehicle | Casualty |
|---|---|---|
| Accident Index | Accident Index | Accident Index |
| Police Force | Vehicle Reference | Vehicle Reference |
| Accident Severity | Vehicle Type | Casualty Reference |
| Number of Vehicles | Towing and Articulation | Casualty Class |
| Number of Casualties | Vehicle Manoeuvre | Sex of Casualty |
| Date (DD/MM/YYYY) | Vehicle Location-Restricted Lane | Age of Casualty |
| Day of Week | Junction Location | Age Band of Casualty |
| Time (HH:MM) | Skidding and Overturning | Casualty Severity |
| Location Easting OSGR (Null if not known) | Hit Object in Carriageway | Pedestrian Location |
| Location Northing OSGR (Null if not known) | Vehicle Leaving Carriageway | Pedestrian Movement |
| Longitude (Null if not known) | Hit Object off Carriageway | Car Passenger |
| Latitude (Null if not known) | 1st Point of Impact | Bus or Coach Passenger |
| Local Authority (District) | Was Vehicle Left Hand Drive | Pedestrian Road Maintenance Worker |
| Local Authority (Highway Authority - ONS code) | Journey Purpose of Driver | Casualty Type |
| 1st Road Class | Sex of Driver | Casualty IMD Decile |
| 1st Road Number | Age of Driver | Casualty Home Area Type |
| Road Type | Age Band of Driver | |
| Speed limit | Engine Capacity | |
| Junction Detail | Vehicle Propulsion Code | |
| Junction Control | Age of Vehicle (manufacture) | |
| 2nd Road Class | Driver IMD Decile | |
| 2nd Road Number | Driver Home Area Type | |
| 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 | | |
| Lower Super Ouput Area of Accident_Location (England & Wales only) | | |

*Figure 1: Data - List of Key variables*

5) Data summary: Each row in the dataset essentially describes the circumstances of personal injury road accidents in the UK – the types of vehicles involved and the consequential casualties. The dataset is limited to include cases that were reported to the police and subsequently recorded using the STATS19 accident reporting form **(assets.publishing.service.gov.uk, 2020)**. Thereby, damage-only accidents with no human casualties or accidents on private roads or car parks are not included in the dataset. Death figures are indicative of persons killed immediately or who died within 30 days of the accident. (Definition adopted by the Vienna Convention in 1968).

## 3. Methodology

## 4. Results

## 5. Discussion

## 6. Conclusion

# Bibliography

assets.publishing.service.gov.uk, 2020. *ACCIDENT STATISTICS Other ref.*. [Online]
Available at:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachm
ent_data/file/230590/stats19.pdf
[Accessed 11 October 2020].

data.gov.uk, 2020. *Road Safety Data*. [Online]
Available at: https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-
47e5ce24a11f/road-safety-data
[Accessed 11 October 2020].