



Capstone Project

**CAR ACCIDENT
SEVERITY**



CAR ACCIDENT SEVERITY

01 Introduction

In this project, I am going to explore, use and create a model based on the Road Accident Safety Data provided by the UK government for the year of 2018. Because it is a labeled dataset, I am going to use Supervised Machine Learning. Also this work is based on the Cross-industry standard process for data mining (CRISP-DM) and its six stages:

1. Business Understanding
2. Data Understanding
3. Data Preparation and Cleaning
4. Modeling
5. Evaluation
6. Deployment



CAR ACCIDENT SEVERITY

02 Business Problem

Engineers and researchers in the automotive industry have tried to design and build safer cars, but road accidents are inevitable. The large number of people, means of transport and different climatic or circumstantial factors have made traffic accidents a part of everyday life around the world. The patterns involved in dangerous crashes and road accidents can be helpful in developing road safety policies or greater awareness of the driver himself. The study of these patterns, causes and severity of injuries can predict the severity or probability of an accident, which would be very useful for the well-being of any driver, offering the possibility that, knowing the risks and probabilities of an accident, people drive with more care. or even change their trip. Undoubtedly, this would help all drivers to have a better perspective of possible accidents, the most dangerous routes, the less favorable weather conditions and with all these data, to make a more informed decision.



CAR ACCIDENT SEVERITY

03 Data

For the analysis and the creation of a predictive model and because the dataset is labeled, I am going to use supervised machine learning. The target variable to predict will be **Accident Severity** which defines three modes: Fatal, serious, and light.

First, the information will be analyzed to determine which of the attributes have the highest correlation with the target variable, among which are probably: Light Conditions, Weather Conditions, Road Surface Conditions, Junction Detail, and some others. Once identified, the data will be cleaned up based on the data visualization for modeling.

Important Data Features

The dataset variables will be grouped into three main sets:

4

Accident details

- 1.Accident Index
- 2.Police Force
- 3.Accident Severity
- 4.Number of Vehicles
- 5.Number of Casualties
- 6.Did Police Officer Attend Scene of Accident
- 7.1st Road Class
- 8.1st Road Number
- 9.Road Type
- 10.Speed limit
- 11.Junction Detail
- 12.Junction Control
- 13.2nd Road Class
- 14.2nd Road Number
- 15.Pedestrian Crossing-Human Control
- 16.Pedestrian Crossing-Physical Facilities

Location and Time

- 1.Location Easting OSGR (Null if not known)
- 2.Location Northing OSGR (Null if not known)
- 3.Longitude (Null if not known)
- 4.Latitude (Null if not known)
- 5.Date (DD/MM/YYYY)
- 6.Day of Week
- 7.Time (HH:MM)
- 8.Local Authority (District)
- 9.Local Authority (Highway Authority - ONS code)
- 10.Urban or Rural Area

Environmental Issues

- 1.Carriageway Hazards
- 2.Light Conditions
- 3.Weather Conditions
- 4.Road Surface Conditions
- 5.Special Conditions at Site

CAR ACCIDENT SEVERITY

04 General Approach

With all the data above, it is possible to determine interesting situations and make questions about them.

1. When do most accidents happen? On weekdays or the weekend?
2. Is weather a key factor in road accidents?
3. Is the number of people affected related to the severity of the accident?
4. Which causes more accidents?



CAR ACCIDENT SEVERITY

05 Final Objective

The main objective is to identify the severity or probability of a possible accident, so this work can be useful for anyone interested in assessing risks while driving, the traffic control departments or simply for anyone who drives a vehicle.

