Machine Learning Assignment

Project Proposal

Lewis Quick - 22016949, William Forber - 22015706, Yamin Shwe Yi Htay – 23019880

# Problem

The problem that we are going to investigate is road traffic accidents in the Bristol area. This data only covers road accidents which involved personal injury, therefore excluding incidents which only produced vehicular damage.

# Context and Background

This is an interesting topic to cover as we could be able to use machine learning to highlight reoccurring features of accidents. Understanding these features could provide insight into common circumstances under which accidents occur. Using this information, research could be done into reducing the relevant risk factors, and drivers could also be advised on when to avoid the roads if possible. In general, our research could be used to reduce harm done in traffic accidents.

# Dataset

The dataset is provided by Bristol Council on their open data website (Open Data Bristol, 2017). Data follows the following structure:

|  |  |
| --- | --- |
| Field | Description |
| Date | Date of accident. |
| Time | Time of accident. |
| Severity | Integer value of severity.  (3, 2, 1). |
| Severity Description | Label of severity.  (Slight, severe, fatal). |
| Accident Type | Accident code.  (LC, A, HO, …) |
| Accident Description | Description of accident code.  (Loss of Control, Adult Pedestrian, Head On, …). |
| Vehicles | Number of vehicles involved. |
| Casualties | The number of casualties involved. |
| Pedestrian | The number of pedestrians involved. |
| Cycles | Number of cycles involved. |
| MCycles | The number of motorcycles involved. |
| Children | Number of children involved. |
| OAPs | Number of OAPs involved. |
| X | X coordinates of location. |
| Y | Y coordinates of location. |
| Render | The main cause of the accident.  (Cars, Cyc, A, …)  {Cars, Cycles, Adult pedestrian, …}  e.g. Many different accident types can be attributed to ‘Cars’. |

# Methodology

We will apply multiple classification algorithms being support vector machines (using linear, polynomial and radial basis function kernels), and ensemble methods like bagging and random forest classifiers). Trying different hyperparameters and evaluating the performance of each of these algorithms will allow us to select the best algorithm for the job.

# Evaluation

We will present our findings using both visual and numerical data. Examples include 2D and 3D graphs, as well as numerical statistics presented in clear tables. We will also evaluate the effectiveness of our algorithms by analysing the accuracy and precision of the results. This will involve considering the likes of classification error, and how well the algorithm performs on new datasets (robustness). This is essential for proving a successful machine learning solution.

# References

Open Data Bristol (2017) Traffic Accidents. Available from: https://opendata.bristol.gov.uk/datasets/bcc::traffic-accidents-1/about [Accessed 05/03/2024].