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 various classification algorithms suitable for the problem, including;

* Support Vector Machines (SVM) with linear, polynomial, and radial basis function (RBF) kernels. To implement SVM, we will utilize the ‘**SVC**’ class from scikit-learn, allowing us to train SVM classifiers with different kernel types.
* Ensemble methods such as bagging and random forest classifiers. For bagging, we will employ the ‘**BaggingClassifier**’ class in scikit-learn, enabling us to train an ensemble classifier by combining predictions from multiple SVM classifiers trained on different subsets of data. For random forest classifiers, we will use the ‘**RandomForestClassifier**’ class to build a random forest classifier by aggregating predictions from multiple decision trees. In terms of improving these established implementations, we will customize them as needed by adjusting hyperparameters, experimenting with different functions, or incorporating domain-specific knowledge. By integrating all these algorithms and their implementations, we can deploy a diverse set of classification techniques to predict road traffic accidents in the Bristol Area.

# Evaluation

We will evaluate our predictive models using both qualitative and quantitative methods.

* Visualizations: We will use 2D and 3D graphs , map, and clear tables to illustrate patterns and trends in data.
* Interpretation: We will analyze these visualizations to understand the model behaviour and performance, identifying areas of high accident frequency and trends.
* Performance Metrics: We will assess accuracy , precision, F1 score, recall, and classification error to comprehensively measure model performance.
* Cross-Validation: We will use cross-validation to ensure the robustness of our models and their ability to generalize to new datasets.

By combining visualizations with quantitative analysis, we intend to provide a precise evaluation of our model’s effectiveness .

.

# References

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