

# 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, which could allow for research into reducing the risk factors. 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). It is updated regularly, as recently as 2024, and includes a comprehensive collection of data including the groups involved, and the time/location of the accident.

## 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].