Team 6

Harshada Jivane

Manasa Hari

Pragati Sharma

**Project Idea**

Build a predictive model to predict number of future accidents using time series forecasting.

A classification model to predict the probability whether the route is accident prone given the data of circumstances such as vehicle type, weather, road conditions.

**Proposed Tasks**

1. Basic data mining tasks for data preprocessing, data analysis and feature engineering
2. For time series forecasting, we might use either Prophet (Time Series forecasting algorithm from Facebook) or ARIMA (Auto Regressive Integrated Moving Average) based on which model gives better performance.
3. Classification model to predict whether the route is accident prone

**Dataset**

The dataset consists of accident and vehicle information for the year 2019

There are three different CSV files with following number of records and features:

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Accidents | Casualties | Vehicles |
| 2019 | 216382 | 153159 | 65536 |
| 2018 | 122636 | 160598 | 226410 |
| 2017 | 129983 | 170994 | 238927 |
|  | ***Total Records = 1484625*** | ***Features (23+16+32) = 71*** |  |

The features in these three files have been referenced using a common “Accident\_Index” column. The dataset also contains a variable look-up excel sheet which provides information about various features in the three CSVs.

The dataset is available on UK’s government website: <https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data>