Project Description:

This data set is collected from Addis Ababa Sub-city police departments for master's research work. The data set has been prepared from manual records of road traffic accidents of the year 2017-20. All the sensitive information has been excluded during data encoding and finally, it has 32 features and 12316 instances of the accident. Then it is preprocessed to identify the major causes of the accident by analyzing it using different machine-learning classification algorithms.

Problem Statement:

The target feature is **Accident_severity** which is a multi-class variable. The task is to classify this variable based on the other 31 features step-by-step by going through each day's task. Your metric for evaluation will be **f1-score**

Key Objectives of The Project:

1. Exploratory data analysis

- Data analysis using `dabl`
- Exploratory data analysis using `matplotlib` and `seaborn`

2. Data preparation and pre-processing

- Missing Values Treatment using the 'fillina' method
- One Hot encoding using pandas get_dummies
- Feature selection using `chi2` statistic and SelectKBest method
- Use PCA to reduce dimensionality
- Imbalance data treatment using `SMOTENC` technique

3. Modelling using sci-kit learn library

- Baseline model using `RandomForest` using default technique
- Tuned hyperparameters using `n_estimators` and `max_depth` parameters

4. Evaluation

- The evaluation metric - weighted `f1_score`
- Baseline model evaluation `fl_score`
- Final model evaluation