Project Documentation for Road Traffic Accident Analysis and Prediction

By Unwana Ita

---

1. Project Overview

This project aims to build on previous exploratory data analysis (EDA) of road traffic accidents by investigating accident severity, geographical distribution, data quality, and predictive modeling. The goal is to develop actionable insights and predictive models that can help improve road safety measures.

The previous work on this project can be found here: https://github.com/ekeng3/ACME-software-Lab

2. Objectives

The project is structured around the following key objectives:

- **Uncover Relationships:** Use the Apriori algorithm to identify relationships between various factors influencing road traffic accidents.
- **Geographical Analysis:** Identify and analyze accidents in specific regions by mapping accident data to exact locations using LSOA (Lower Super Output Area) codes.
- **Outlier Detection:** Implement outlier detection methods to identify unusual or suspicious entries in the dataset.
- **Predictive Modeling:** Develop a classification model to accurately predict fatal injuries in road traffic accidents, thereby providing insights for enhancing road safety.

3. Data Processing

- **Region Mapping:** The LSOA codes from the dataset are used to extract and map accident locations to specific regions in the UK. This step is essential for the geographical distribution analysis.
- **Data Cleaning:** The dataset undergoes cleaning and preprocessing to ensure data quality before further analyses and models are applied.

4. Methodology

- **Apriori Algorithm:** The Apriori algorithm is employed to uncover patterns and relationships between various variables related to road traffic accidents.
- **Outlier Detection:** Outlier detection techniques are applied to identify and handle anomalies in the dataset that could affect the analysis.
- **Classification Model:** A machine learning classification model is developed using the processed data to predict the severity of injuries sustained in road accidents.

5. Results

- **Patterns and Relationships:** The Apriori algorithm results in a set of association rules that highlight significant relationships between factors like time of day, weather conditions, and accident severity.
- **Geographical Insights:** The mapping of accidents to specific regions reveals hotspots with a high frequency of severe accidents, which can guide targeted safety interventions.
- **Model Performance:** The classification model demonstrates its ability to accurately predict fatal injuries, with performance metrics such as accuracy, precision, recall, and F1-score used to evaluate the model.

6. Conclusion

The project successfully extends the EDA by introducing advanced analytical methods to gain deeper insights into road traffic accidents. The predictive model developed in this phase provides a valuable tool for policymakers and safety professionals to anticipate and mitigate severe road accidents.

7. Future Work

- **Model Refinement:** Further refinement of the classification model by exploring additional features and advanced machine-learning techniques.
- **Real-Time Analysis:** Implementation of real-time data feeds to continuously update the model and provide live insights.
- **Broader Application:** Extending the analysis to include more comprehensive datasets covering different regions or countries.

Please check out the full work at this GitHub repository:
https://github.com/ekeng3/ACME_Capstone-Project