

ACME SOFTWARE LABS

Final Capstone Project

Project Overview

Building upon the completed exploratory data analysis (EDA), this phase deepens the investigation into road traffic accidents, focusing on accident severity, geographical distribution, data quality, and predictive modeling.

Note: the deadline for the submission of this task **Wednesday, 7th August 2024**.

Project Objectives

- Using the apriori algorithm to uncover relationships between variables influencing accident severity.
- Identify accidents in the regions: Kingston upon Hull, Humberside, and the East Riding of Yorkshire, etc. You can do this by filtering on LSOA, or police region or another method if you can find one. Run clustering on this data. What do these clusters reveal about the distribution of accidents in our region?
- Using outlier detection methods, identify unusual entries in your data set. Should you keep these entries in your data?
- Develop a classification model using the provided data that accurately predicts fatal injuries sustained in road traffic accidents, with the aim of informing and improving road safety measures?

Deliverables

1. Apriori Algorithm Analysis:

- Identification of key variables impacting accident severity.
- Detailed analysis of association rules generated by the apriori algorithm.
- Visualization of significant patterns and trends.

2. Spatial Analysis and Clustering:

- Defined clusters of accident hotspots within the region (Kingston upon Hull, Humberside, and East Riding of Yorkshire).
- Characterization of identified clusters based on relevant attributes.

- Visual representation of accident distribution and clusters (maps, charts).

3. Outlier Detection and Assessment:

- Application of appropriate outlier detection methods.
- Identification and documentation of unusual data points.
- Justification for retaining or removing outliers based on analysis.

4. Predictive Model Development:

- Selection and preparation of relevant features for model training.
- Development and training of a classification model to predict fatal injuries.
- Evaluation of model performance using appropriate metrics (accuracy, precision, recall, F1-score).

5. Final Report:

- Comprehensive summary of findings from all project phases.
- Clear and concise presentation of results, including visualizations.
- Recommendations for road safety improvements based on model insights.