From Data to Insights: A Comprehensive Analysis of Traffic Accident Data in the UK (2000-2022)

Louise Malm[1], Nikolay Moshenskiy[1]

[1]Institute of Computer Science, University of Tartu, Tartu

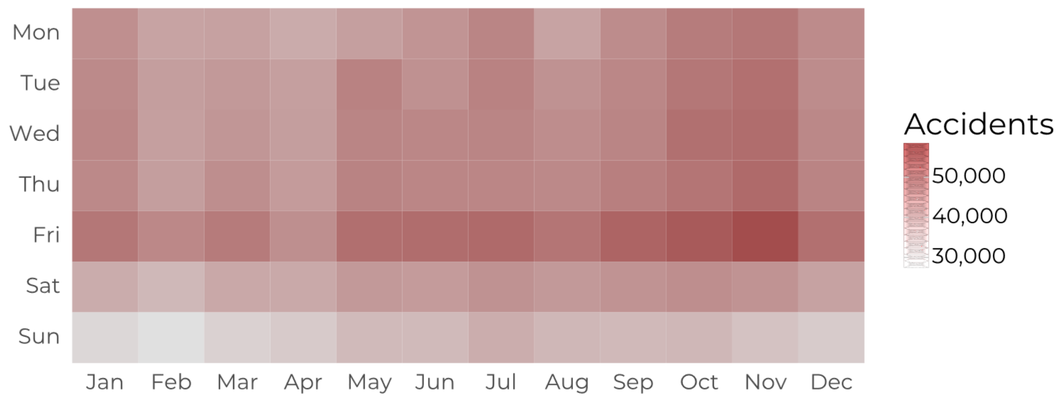
DATA SOURCE

Introduction

In the UK, road accidents cause several hundreds of deaths and tens of thousands of severe injuries annually. Can these numbers, especially those related to fatality, be reduced by insights learned from historical data?

The analysis aims to uncover patterns, trends, and factors contributing to road accidents, especially those leading to fatal outcomes. In addition to analysis, a tool classifying an accident as fatal or non-fatal is expected to provide additional insights into the most significant features contributing to fatal accidents.

Fig. 1. Heatmap of accidents by weekday and month



Data and methodology

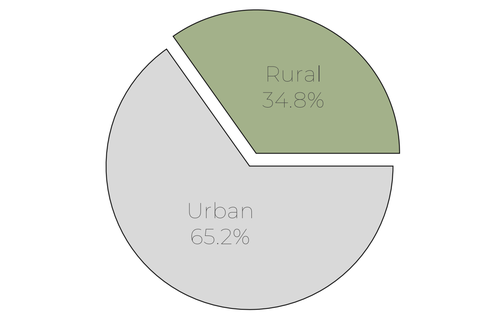
- UK road accident data excluding northern Ireland (2000-2022) detailing collisions and casualties.
- Includes specifics on 3,691,651 collisions and the 4,955,958 casualties involved.

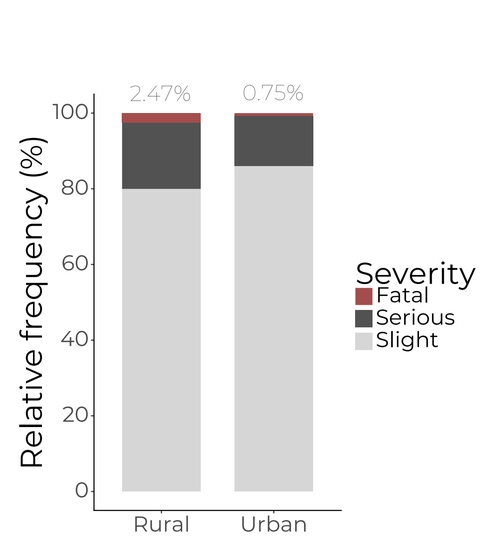
Preparation for machine learning included removing unknown values (leaving 640,359 accidents), one-hot encoding and data balancing. Accident severity was predicted using Random Forest and XGBoost classifiers and evaluated based on AUC and confusion matrix.
Visual analyses include temporal trends and spatial clustering to identify high-risk areas for fatal accidents.

Results

- Number of accidents have steadily decreased during the past decades.
- Most accidents occurred in November on Fridays (Fig. 1) and during morning and evening rush hours.
- Even though more accidents happened in densely populated areas, fatality of the accidents were higher in rural areas (Fig. 2).
- Best classification results obtained from the XGBoost model trained on under sampled data (Table 1 for test results on 96,054 accidents).
- Most influential features for machine learning was junction detail, junction control and speed limit (Fig. 3).

Fig. 2. Road accident distribution in the UK across urban and rural areas





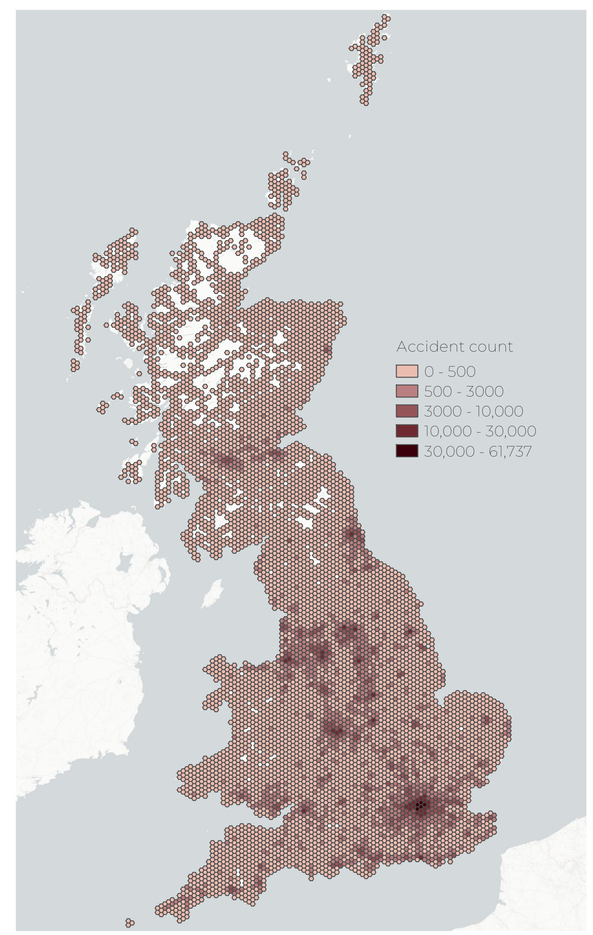


Table 1. Summary of XGBoost classifier

	Non-fatal	Fatal	
Non-fatal	TN: 69,536	FP: 25,606	AUC
Fatal	FN: 316	TP: 596	74.86%

Fig 3. Accident count with fatal percentage vs (A) junction detail, (B) junction control, and (C) speed limit