An Analysis of Contributing Factors in Severe and Fatal Road Traffic Accidents in the UK

2022/2023 CA683I Data Analytics and Data Mining

Project Group 30 Notebook: https://github.com/te-dcu/CA683I-2023-Assignment/Notebook.ipynb

Nuno Correia
#21267090
School of Computing
Dublin City University
nuno.correia2@mail.dcu.ie

Tristan Everitt
#22270316
School of Computing
Dublin City University
tristan.everitt2@mail.dcu.ie

Toyatma Fedee
#22267485
School of Computing
Dublin City University
toyatma.fedee2@mail.dcu.ie

Paul Ryan
#22270321
School of Computing
Dublin City University
paul.ryan79@mail.dcu.ie

***Abstract*—This study examined UK road traffic accidents between 2005 and 2017, focusing on determining the factors significantly related to accident severity. Out of a total of 2,058,408 incidents, 14% were classified as severe accidents. The study employed a Chi-square test to analyze the relationship between 219 features and accident severity, discovering that 198 features rejected the null hypothesis, thus indicating a significant relationship with accident severity.**

The most influential factors included vehicle type (motorcycles being particularly prominent), speed limits (specifically 60 miles per hour), the first point of impact, various vehicle manoeuvres (such as waiting to go or slowing down), light conditions (darkness with no lighting), vehicles leaving the carriageway (either nearside or offside), hitting objects off the carriageway (notably trees), and junction location (not at or within 20 meters of a junction). It is important to note that these factors do not imply causation but rather a higher prevalence of severe accidents compared to less severe ones.

To optimise the prediction of accident severity using a machine learning model, Recursive Feature Elimination (RFE) was utilised. This method aimed to identify the ideal balance between the number of features and model performance. The RFE process pinpointed 43 features as the optimal balance, suggesting that incorporating these top factors in the model is likely to yield the most accurate predictions for accident severity.

I. Introduction

The World Health Organisation released a sobering report at the end of 2018, and it highlighted that road traffic deaths has reached 1.35 million per year [1]. Road traffic related deaths have become the main cause of death of people aged between 5 and 29 years and the eight leading cause for all age groups, surpassing HIV/AIDS [1]. The UK alone reported an estimated 1,560 road deaths for 2021 [2]. With people travelling more miles in Great Britain increasing year-on-year, accidents and their prevention become a big concern for policymakers [2][1].

This study seeks to identify factors that contribute to serious road accidents using the Road Traffic Accident Data [3] provided by the UK Department for Transport for the period of 2005–2017 containing 2,058,408 incidents. A serious accident is defined as incidents requiring hospitalisation or resulting in fatalities.

- **The Null hypothesis** ($H0$): There is no significant relationship between the factors under consideration and the severity of accidents.
- **The Alternative hypothesis** ($H1$): There exists a significant relationship between at least one of the factors under consideration and the severity of accidents.

In order to assess the significance of the factors, we will use $\alpha = 0.05$ for the null hypothesis ($H0$). This α value represents a 95% level of confidence that we wish to achieve before a factor rejects the null hypothesis.

By exploring the relationships between the factors and accident severity, the study aims to contribute to existing knowledge and provide insights that can help enhance road safety measures and potentially save lives.

Section II examines existing literature on different road accident risk factors and the methods used for analysing them. Section III outlines the methodology employed by this paper. Section IV presents the findings, while section V offers concluding remarks and future work.

II. Related Work

A review of existing literature was conducted to observe different methods and techniques that were applied to study factors that increase the risk of serious road crashes.

"Risk factors for fatal road traffic accidents in Udine, Italy" [4] utilised data from regional authorities and employed statistical analysis techniques to identify significant factors that contributed to fatal road accidents. Lack of wearing a seatbelt, speeding, age cohort, poor visibility were found as contributing factors. The study used logistic regression to evaluate a connection between driver attributes and the circumstances of severe road traffic accidents. However, the paper acknowledged that there were limitations in the data

[1]Due to the coronavirus (COVID-19) pandemic, the report's long term trends for 2021 can be misleading.

that was collected, in particular information about seat belt usage and alcohol and drug usage.

The next paper reviewed used a similar dataset that this paper used. "Association Rule Mining for Road Traffic Accident Analysis: A Case Study from UK" [5] applied the Apriori algorithm used in Association Rule Mining approach along with network visualisation to analyse UK road accident data from 2005 to 2017. Potential further studies regarding the method used by the paper can look closer to see if the rules change significantly between cities and also how it would apply to crime.

"Exploring the factors affecting motorway accident severity in England using the generalised ordered logistic regression model" [6] utilised UK road traffic accident data between 2005 and 2011, but with a focus on accidents that took place on the hard-shoulder of the motorways. The study used Generalised Ordered Logistic Regression (logit) to analyse how certain factors influenced the likelihood of three ranked accidents (Fatal, Severe, Slight) occurring. Factors found to increase the severity of the accident involved the hour of the day when there is high levels of traffic, visibility, and fatigue in Heavy Goods Vehicle (HGV) drivers.

"A data mining framework to analyze road accident data" [7] examined road accidents in a small Indian region, proposing a two-phase process using K-Modes clustering and association rule mining for each cluster and the entire dataset. They identified six distinct clusters and their specific characteristics.

"Road Accident Analysis with Data Mining Approach: evidence from Rome" [8] analysed Rome Municipality road accident data using K-Means clustering, Kohonen networks, decision trees, association rule mining, and artificial neural networks, concluding that vehicle type significantly influences accident severity.

"Road Traffic Accidents Injury Data Analytics" [9] looks at data from a similar source as we do and for a similar period. It combines three elements - Accident information, Vehicle information and Casualty information. Interestingly, many categorical features with multiple possible values were reduced to binary variables. The paper examines the use of XGBoost the most accurate at 74.4% to determine what are the key factors contributing to severity. They concluded that Casualty Type has the greatest bearing on accident severity.

"Analyzing Factors Associated with Fatal Road Crashes: A Machine Learning Approach" [10] examines the problem of fatal road collisions. Using machine learning methods, the study tries to uncover the variables related with deadly road collisions. The study looked at the relationship between road features, driver-related variables, vehicle types, and meteorological conditions and the occurrence of fatal road collisions. The outcomes of the research demonstrated that various variables, including driver-related factors such as speeding, distracted driving, and driving while impaired by drugs or alcohol, were substantially correlated with fatal road collisions.

"What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records" [11] employed a mixed-methods approach to study road accidents, revealing key contributing factors despite the subjective nature of survey and interview data. The study underscores the importance of effective road safety measures and implementing appropriate data mining methods like association rule mining.

"Comparison of Statistical and Machine-Learning Models on Road Traffic Accident Severity Classification" [12] compares the effectiveness of several statistical and machine learning models for predicting the severity of traffic accidents. The study emphasises the need for precise accident severity categorisation to help improve road safety and reduce deaths. Machine learning approaches outperformed traditional statistical models in predicting accident severity.

"Analysis of road traffic fatal accidents using data mining techniques" [13] utilised data processing techniques to analyse a large dataset concerning road accidents. They examined the correlation between fatality rates and various factors, including collision types, weather conditions, road surface conditions, lighting conditions, and the presence of intoxicated drivers. They employed the Apriori algorithm to identify association rules, designed a classification model using the Naive Bayes classifier, and created clusters using the simple K-means clustering method.

"Analysis of road accidents using data mining techniques" [14] extracted frequent road accident patterns using association and classification rules. They used the Apriori algorithm to analyse historical data and predict accident types on both existing and new roads.

"Identifying Efficient Road Safety Prediction Model Using Data Mining Classifiers" [15] studied road accidents with data mining techniques, employing algorithms like Naive Bayes, Random Forest, and J48. Their analysis aimed to uncover incident reasons, pinpoint high-risk areas, and identify leading causes of severe accidents.

III. Methodology

A knowledge discovery structure was carried out to standardise and arrange the knowledge discovery components of the study. The Knowledge Discovery in Databases (KDD) process was utilised, in conjunction with the scikit-learn library [16] for data collection, preprocessing, dimensionality reduction, and algorithm implementation. KDD is a five-stage[2] iterative process that outlines the data analysis life-cycle.

A. Data Selection and Exploratory Analysis

The source of the data set was obtained from Kaggle [17], which is derived from the United Kingdom Department for Transport *Road Traffic Accident Custom Downloads* [18]. The dataset contains accident data reporting to police where at least one person was injured and comes in two parts:

- Accident data
- Vehicle data

We joined the two datasets on an `Accident_Index` identifier.

[2]Some stages can be further split thus making a seven-stage process.

B. Data Pre-Processing

First of all we decided to remove records from 2004 as Vehicle information is missing from those records. The dataset contains spatial coordinates in the form of Latitude and Longitude, and was subsequently used to inform the process of filling other missing data where appropriate. There were a small number of missing Latitude/Longitude missing values to begin with - these were filled using the mode of the containing Local Area District. In this way we maintained some element of location. Once we had Latitude/Longitude complete we built a Nearest Neighbour structure using the entire dataset. Missing values for the columns LSOA, Road Number, Speed Limit, Pedestrian Crossing and if in Scotland were imputed by querying the `NearestNeighbour` structure for the nearest spatial neighbour(s) and setting the missing value based on that information. We used `sklearn.neighbours.BallTree` [16] to create the `NearestNeighbour` structure based on latitude and longitude of all data (see Fig. 1).

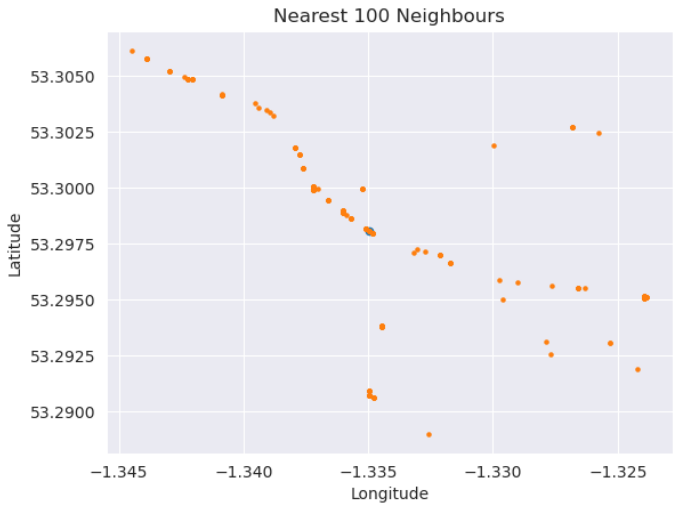

Fig. 1. BallTree NearestNeighbour structure based on latitude and longitude

While this is slower than using mode or median, it does lead to far more accurate data. Missing values for vehicle information was based on a loose hierarchy of Vehicle Type, Make, Model and Engine Capacity. So for instance, when populating missing values for Model, we grouped by Make and used the mode of the group rather than the mode of the entire column. To reduce the dataset and eliminate redundant variables, the following were removed:

- Latitude/Longitude was used instead to fill missing data:
  - Location Easting (OSGR)
  - Location Northing (OSGR)
  - Local Authority (District)
  - Local Authority (Highway)
- Dependent variables on top of the Accident Severity:
  - Did Police Officer Attend Scene of Accident
  - Number of Casualties
- Irrelevant to our goal:
  - Vehicle Reference
  - LSOA of Accident Location
  - Police Force
  - In Scotland
  - 1st Road Number
  - 2nd Road Number
  - Year
- Related to other variables or made redundant after pre-processing:
  - Time
  - Latitude/Longitude
  - Urban or Rural Area

C. Data Transformation

- Numeric columns are standardised by scaling them between 0 and 1.
- Columns with only whole numbers converted from decimal to integer.
- `Accident_Severity` is the dependent variable in our analysis, therefore Fatal and Serious were encoded as 1 and Slight to 0.
- Hour of Day $\rightarrow$ Early Morning, Late Morning, Early Afternoon, Late Afternoon, Evening, Night.
- Date $\rightarrow$ Day of Month, Day of Year
- Vehicle Type/Model and Engine Capacity $\rightarrow$ Commercial or Public Transport, Motorcycle, Regular Car, Sport Car, Large Engine Capacity Car, Small Engine Capacity Car, Other.
- And the remaining categorical columns were One-Hot encoded.

The data transformation process resulted in a total of 219 features for us to then analyse further.

D. Data Mining

We conducted a Chi-square test on each feature to assess which ones reject or fail to reject the null hypothesis ($H0$). We conduced a Ch-square test to identify which features (e.g., road conditions, weather, vehicle type, etc.) have a significant relationship with accident severity. The test uses a significance level of $\alpha = 0.05$.

If the p-value is less than or equal to 0.05, it is concluded that there is a significant relationship between the feature and accident severity. The critical value is another approach to deciding about the null hypothesis. If the Chi-square statistic is greater than the critical value, the null hypothesis is rejected, indicating that the feature has a significant impact on accident severity.

Each feature is analysed, and its Chi-square statistic, p-value, degrees of freedom, and critical value are calculated. Based on these values, the features are classified into two groups: those with a significant relationship to accident severity and those without a significant relationship. This classification is done using both the critical value and the p-value methods, but either one works since both should match up.

We also used the Recursive Feature Elimination (RFE) technique to identify which features are more relevant for

predicting the severity of an accident. Computing the ANOVA F-value was considered initially, but computing Chi-square statistics was preferred due to the nature of the binary input. The process was to split the data into training and testing and use the best k features for predicting our target class. The value k started by including all of our features and iteratively reduce k by one for each prediction. The Bernoulli Naive Bayes classifier was used since it is fast and suitable for one-hot encoded data and the recall value was recorded. The recall value was preferred over the F1, accuracy and precision since we wanted to ensure that false negatives would be minimised. Recall measures the proportion of actual positive outcomes (severe/fatal) that are correctly identified by our classifier and helps us capture a larger percentage of the true positive outcomes.

IV. Evaluation/Results

During our initial investigation, we identified a total of 2,058,408 road traffic incidents in the UK between 2005 and 2017. Of these, 292,758 were categorised as severe accidents, accounting for 14% of the overall incidents. We began with 55 features to examine, and after refining the data by removing irrelevant factors and creating new ones, we were left with 219 features to analyse.

1) Inferential Statistical Test: We conducted a Chi-square test on each feature to assess which ones reject or fail to reject the null hypothesis ($H0$). There are two methods to evaluate the null hypothesis using the Chi-square test: the test statistic method and the p-value method. The test statistic method compares the calculated test statistic against a critical value from the Chi-square distribution table. The p-value method compares the calculated p-value against a significance value, which we defined as $\alpha = 0.05$.

Both methods were examined as an extra check since both should produce the same results.

Out of the 219 features, only 21 failed to reject $H0$ (See Fig. 3). Many features that rejected $H0$ had p-values near zero and a high Chi-square statistic (See Fig. 2), indicating a significant relationship between the factor and accident severity. As our objective was to find at least one factor that could reject $H0$, we can conclude that $H0$ is rejected, and the alternative hypothesis ($H1$) is accepted.

From the 198 that reject $H0$, the following are the 10 showing the most significant relationship.

- Vehicle Type: Motorcycle
- Speed Limit: 60 miles per hour
- 1st Point of Impact: Back
- Vehicle Manoeuvre
  - Waiting to go, held up
  - Slowing or stopping
- Light Conditions: Darkness, no lighting
- Vehicle Leaving Carriageway
  - Nearside
  - Offside
- Hit Object off Carriageway: Tree

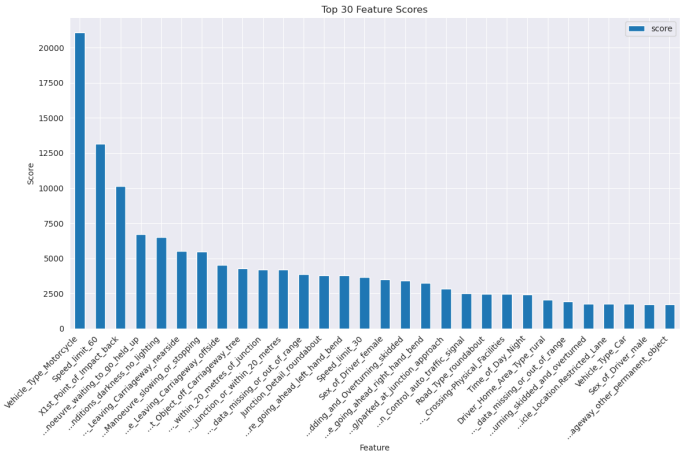

Fig. 2. Chi-squared test using both categorical and numeric attributes.

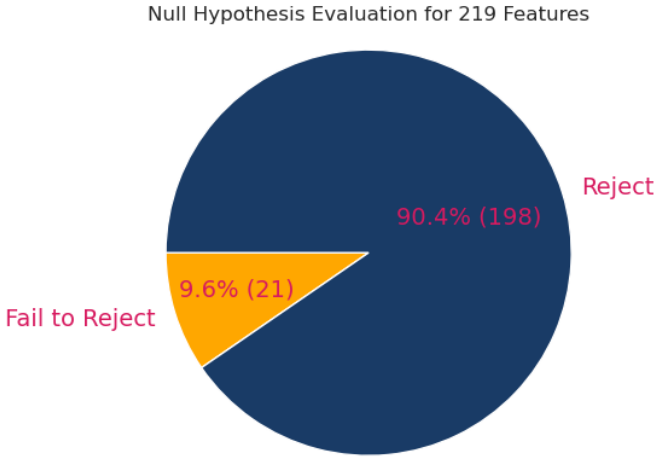

Fig. 3. Chi-squared Test: Null Hypothesis $H0$ Evaluation.

- Junction Location
  - Not at or within 20 metres of junction

However, only because a feature has a significant relationship with a severe accident, it does not imply causation such as traveling at 60 miles per hour will necessarily result in a hospital visit or death. What the results indicate is the prevalence in a severe accident compared to less severe accidents.

Table I presents the total number of incidents for each factor listed above, as well as the percentage contribution to the entire dataset and exclusively to severe incidents and Table II are the factors that fail to reject $H0$.

2) Dimensionality Reduction: We used the Recursive Feature Elimination method to determine the optimal number of features for predicting the severity of road traffic accidents using a machine learning model. It starts with all available features and then iteratively removes one feature at a time by conducting a Chi-squared test to measure the strength of the relationship between each feature and the target variable. The

	Incidents	% All	% Severe
Vehicle Type: Motorcycle	173,110	8.41%	2.42%
Speed Limit: 60	317,097	15.41%	3.51%
1st Point of Impact: Back	388,912	18.89%	1.41%
Waiting to Go / Held Up	145,922	7.09%	0.38%
Slowing or Stopping	171,469	8.33%	0.56%
Darkness, No Lighting	99,708	4.84%	1.21%
Leaving Carriageway: Nearside	123,874	6.02%	1.39%
Leaving Carriageway: Offside	63,189	3.07%	0.78%
Hit Object off Carriageway: Tree	27,633	1.34%	0.41%
Not at or within 20m of Junction	808,108	39.26%	6.76%

TABLE I
SIGNIFICANT ASSOCIATIONS WITH SEVERE ACCIDENTS

	Incidents	% All	% Severe
Hit object: Bridge (roof)	368	0.18%	0.03%
Speed limit 15	11	0.01%	0.00%
Defective road sign or marking	3,197	15.53%	2.22%
Age band of driver 0–5	125	0.61%	0.09%
Special condition: Mud	5,193	25.23%	3.63%
Commuting to/from work	202,525	98.39%	14.03%
Left-hand drive vehicle	3,887	18.88%	2.60%
Age band of driver 6–10	882	4.28%	0.66%
Carriageway hazard: Vehicle load	2,852	13.86%	2.06%
Speed limit 10	13	0.06%	0.02%
Snowing with high winds	2,635	12.80%	1.70%
Speed limit 40	187,000	90.85%	12.81%
Overtaking nearside	10,658	51.78%	7.65%
Speed limit 70	183,791	89.29%	12.82%
Journey as part of work	391,713	190.30%	26.89%
Raining with high winds	27,813	13.51%	1.87%
Hit object: Road works	889	4.32%	0.71%

TABLE II
FEATURES THAT FAIL TO REJECT $H0$

model's performance (recall score) on a test set is recorded and then continues with the iterative process until there are no more features to test. The training/test split we used is 80% to 20% ratio.

By plotting the recall scores against the number of features, an `elbow` point can be identified, which represents the ideal balance between the number of features and the model's performance. In this case, the elbow point is at 43 features (See Fig. 4), suggesting that using these top 43 features is likely to yield the best prediction results for accident severity.

V. CONCLUSIONS AND FUTURE WORK

An analysis of the UK road traffic accident dataset from 2005 to 2017 revealed a total of 2,058,408 incidents, with 14% (292,758) categorised as severe accidents. Initially, 55 features were examined, but after refining the dataset, 219 features were left for analysis. We conducted a Chi-square test on each feature to determine which factors were significantly related to accident severity. Out of these, 198 features rejected the null hypothesis, with the top 10 factors being vehicle type (motorcycle), speed limit (60 mph), first point of impact (back), vehicle manoeuvres (waiting to go/held up and slowing or stopping), light conditions (darkness without lighting), vehicle leaving the carriageway (nearside and offside), hitting

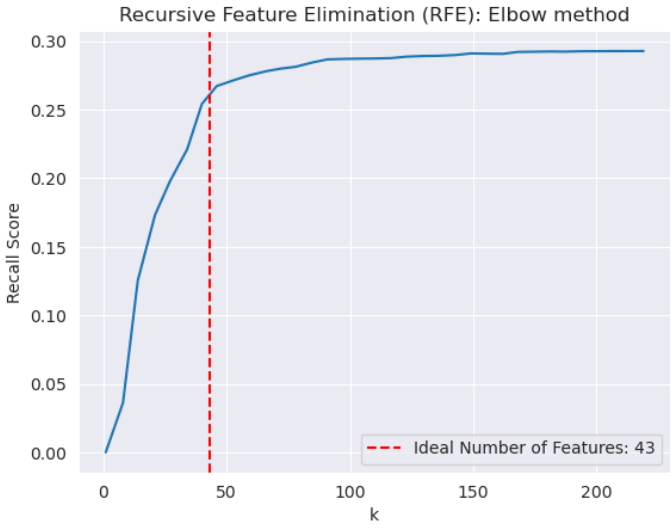

Fig. 4. Recursive Feature Elimination (RFE) Elbow Method

an object off the carriageway (tree), and junction location (not at or within 20 meters of a junction).

We then used Recursive Feature Elimination to find the ideal number of features for predicting accident severity using a machine learning model. This method starts with all available features and iteratively removes one feature at a time, measuring the relationship between each feature and the target variable with a Chi-squared test. The model's performance on a test set is recorded, and the process continues until no more features are left. In this case, the ideal balance between the number of features and the model's performance was found to be 43 features, which is likely to yield the best prediction results for accident severity.

After conducting our analysis, we discovered that there are many avenues to take to get further insights. This study only covered the period between 2005 and 2017, so there is more to explore in the 2018, 2019, and the COVID-19 pandemic years. The United Kingdom Department for Transport also provides more accident attributes that were not included in this study.

At the start of the study, the aim was to obtain all the attributes, but it proved too difficult to obtain them all without having the dataset fail to generate and download. Also, our data transformation effectively aligned with the study's objective, but it may not be ideally suited for machine learning prediction models, as evidenced by our model's limited predictive capabilities (recall score). There were also some unanswered questions, such as how the speed limit of 70 miles per hour failed to reject the null hypothesis, and better ways to clean or reshape the data to eliminate unrealistic values, such as a horse with a value for Engine Capacity. Future work should explore alternative data transformations or structuring approaches to enhance prediction performance.

A potential direction for future work is to apply association rule mining techniques to discover and analyse interesting relationships between contributing factors of road traffic accidents.

Another promising direction for future analysis is to compare the UK road traffic accident dataset with similar datasets from other countries, such as:

- The French Road Safety Observatory (ONISR)
- The Federal Statistical Office of Germany (Statistisches Bundesamt)
- The Australian Bureau of Infrastructure, Transport and Regional Economics (BITRE)
- The Fatality Analysis Reporting System (FARS) provided by the US National Highway Traffic Safety Administration (NHTSA)

By conducting an international comparison, similarities and differences can be further investigated for patterns and contributing factors that might not have been captured in the UK's road traffic accident dataset.

References

[1] W. H. Organization, "Global status report on road safety 2018," World Health Organization, Geneva, Technical Report, 2018. [Online]. Available: https://www.who.int/publications/i/item/9789241565684.

[2] "Road traffic estimates in great britain 2021," Department for Transport, London, Technical Report, 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1107056/road-traffic-estimates-in-great-britain-2021.pdf.

[3] D. for Transport, *Road traffic accident data*, https://roadtraffic.dft.gov.uk/custom-downloads/road-accidents, Accessed: 2023-04-02, 2023.

[4] F. Valent, F. Schiava, C. Savonitto, T. Gallo, S. Brusaferro, and F. Barbone, "Risk factors for fatal road traffic accidents in udine, italy," *Accident Analysis Prevention*, vol. 34, no. 1, pp. 71–84, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457500001044.

[5] M. Feng, J. Zheng, J. Ren, and Y. Xi, "Association rule mining for road traffic accident analysis: A case study from uk," in *Advances in Brain Inspired Cognitive Systems*, J. Ren, A. Hussain, H. Zhao, *et al.*, Eds., Cham: Springer International Publishing, 2020, pp. 520–529.

[6] P. Michalaki, M. A. Quddus, D. Pitfield, and A. Huetson, "Exploring the factors affecting motorway accident severity in england using the generalised ordered logistic regression model," *Journal of Safety Research*, vol. 55, pp. 89–97, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022437515000833.

[7] S. Kumar and D. Toshniwal, "A data mining framework to analyze road accident data," *Journal of Big Data*, vol. 2, no. 26, 2015. [Online]. Available: https://doi.org/10.1186/s40537-015-0035-y.

[8] A. Comi, A. Polimeni, and C. Balsamo, "Road accident analysis with data mining approach: Evidence from rome," *Transportation Research Procedia*, vol. 62, pp. 798–805, 2022, 24th Euro Working Group on Transportation Meeting. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352146522002265.

[9] M. K. Nour, A. Naseer, B. Alkazemi, and M. A. Jamil, "Road traffic accidents injury data analytics," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 12, 2020. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2020.0111287.

[10] A. J. Ghandour, H. Hammoud, and S. Al-Hajj, "Analyzing factors associated with fatal road crashes: A machine learning approach," *International Journal of Environmental Research and Public Health*, vol. 17, no. 11, 2020. [Online]. Available: https://www.mdpi.com/1660-4601/17/11/4111.

[11] J. Rolison, S. Regev, S. Moutari, and A. Feeney, "What are the factors that contribute to road accidents? an assessment of law enforcement views, ordinary drivers' opinions, and road accident records," *Accident Analysis Prevention*, vol. 115, Jun. 2018.

[12] P. Infante, G. Jacinto, A. Afonso, *et al.*, "Comparison of statistical and machine-learning models on road traffic accident severity classification," *Computers*, vol. 11, no. 5, 2022. [Online]. Available: https://www.mdpi.com/2073-431X/11/5/80.

[13] L. Li, S. Shrestha, and G. Hu, "Analysis of road traffic fatal accidents using data mining techniques," in *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, 2017, pp. 363–370.

[14] P. Shetty, S. P. C, S. V. Kashyap, and V. Madi, "Analysis of road accidents using data mining techniques," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 4, pp. 1494–1496, 2017. [Online]. Available: https://www.irjet.net/archives/V4/i4/IRJET-V4I4306.pdf.

[15] D. Karthik, P. Karthikeyan, S. Kalaivani, and K. Vijayarekha, "Identifying efficient road safety prediction model using data mining classifiers," *VOLUME-8 ISSUE-10, AUGUST 2019, REGULAR ISSUE*, 2019.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[17] A. Tsiaras, *Uk road safety: Traffic accidents and vehicles*, 2021. [Online]. Available: https://www.kaggle.com/datasets/tsiaras/uk-road-safety-accidents-and-vehicles.

[18] D. for Transport, *Road traffic accident custom downloads*, https://roadtraffic.dft.gov.uk/custom-downloads/road-accidents, Department for Transport, n.d. [Online]. Available: https://roadtraffic.dft.gov.uk/custom-downloads/road-accidents.