World Academy of Science, Engineering and Technology
International Journal of Transport and Vehicle Engineering
Vol:17, No:7, 2023

Road Safety in Great Britain: An Exploratory Data Analysis

Jatin Kumar Choudhary, Naren Rayala, Abbas Eslami Kiasari, Fahimeh Jafari

***Abstract*—**Great Britain has one of the safest road networks in the world. However, the consequences of any death or serious injury are devastating for loved ones, as well as for those who help the severely injured. This paper aims to analyse Great Britain's road safety situation and show the response measures for areas where the total damage caused by accidents can be significantly and quickly reduced. For the past 30 years, the UK has had a good record in reducing fatalities over the past 30 years, there is still a considerable number of road deaths. The government continues to scale back road deaths empowering responsible road users by identifying and prosecuting the parameters that make the roads less safe. This study represents an exploratory analysis with deep insights which could provide policy makers with invaluable insights into how accidents happen and how they can be mitigated. We use STATS19 data published by the UK government. Since we need more information about locations which is not provided in STATA19, we first expand the features of the dataset using OpenStreetMap and Visual Crossing. This paper also provides a discussion regarding new road safety methods.

***Keywords*—**Road safety, data analysis, OpenStreetMap, feature expanding.

I. Introduction

TRANSPORTATION brings immense benefits to society, but it also has its costs. Costs include those of infrastructure, personnel and equipment, but also the loss of life and property in traffic accidents on the road, delays in travel due to traffic congestion and various indirect costs in terms of air transport.

Road crashes are associated with 1.2 million deaths and 50 million injuries or disabilities each year worldwide [1]. Many governments and non-profit agencies collect driving and riding data to produce many statistics and visual graphs to understand road crash causes and to propose solutions for different roads and driving zones. More research has been done to identify the various factors that affect road accidents such as road infrastructure, traffic and sociodemographic characteristics, land use and the environment.

In previous investigations [2], [3], road casualties were measured at various regional boundaries, such as highways, intersections, section of roads or counties, districts, traffic zones, postcodes, or other census tracts. A common approach is to develop relationships between traffic accidents and contributing factors using different statistical models. This relationship is developed at the level of the road or individual area.

According to road casualty statistics [2], young men between the ages of 17 and 24 years are the highest risk group of all motorists. The current education focus is on friends being vigilant and helping each other to reduce dangerous behaviour behind the wheel of a car. Peer dynamics are critical to young men's decision making. A report from the RAC Foundation [4] reports that chasing a herd is very important for young drivers who may be hesitant to speak up if they feel unsafe because they are afraid of being judged negatively by their friends.

In order to conduct this study, we utilized OpenStreetMap (OSM), a collaborative project which provides free access to crowdsourced spatial data. A country's data representativeness may differ from another because the data are collected by volunteers. As of March 2022, OSM data in Great Britain contain over three million road segments that adequately cover the whole area.

As part of the study, we conducted exploratory data analysis using STATS19 data [5], [6]. In addition, using OSM and Overpass API, we build a map server and then retrieve all road segments and road nodes and dump them to a database of road data. Then we clean the road data and expand the map-based static features. The components of the input data that do not vary over time are referred to as static features. There are elements derived from road geometry, such as geohash, segment curvature, orientation, proximity to nearby intersections, as well as number of amenities and shops around a node [7]. In addition, we use Visual Crossing weather API to incorporate weather data into the dataset.

The remainder of this paper is organized as follows. Section II discusses the related works. Section III introduces the problem statement describing the details regarding data collection and data pre-processing. Section IV provides exploratory data analysis with deep insights. Section V provides a discussion regarding the challenges and opportunities. Section VI represents some of the current solutions for the discussed challenges. Section VII concludes this paper and gives ideas for future works.

II. Related Works

This section reviews some of the relevant papers in road safety. Over the past decades, speed cameras have been used in the UK to improve road safety. Although speed camera safety has been estimated extensively, few studies have been carried out on the factors that should be considered when selecting camera locations. A systematic literature review has been represented in [8] to investigate machine learning applied to

Fahimeh Jafari is with University of East London, United Kingdom (e-mail: f.jafari@uel.ac.uk).

road safety modelling. Authors in [9] evaluate the UK safety performance-based criteria for speed camera site selection. A total of 332 speed cameras and 2,513 control sites with road traffic accident data are monitored between 2002 to 2010. The authors employ and compare the probability score matching method and the empirical Bayes method to estimate the safety effects of speed cameras under various scenarios. Findings indicate that the proximity to residential neighbourhoods and school zones is an important factor for selecting speed camera locations, as is the density of the population. There is also an evaluation of the official criteria used to select camera sites, such as previous accident history, site length, and risk value. For optimum speed camera safety effects, it is suggested to utilize a length of 500 meters for the site.

We will live and work differently as a result of the ongoing fourth industrial revolution. There are some areas where innovation is widespread and visible. The progress in other areas, such as transport safety, is much slower. Authors in [10] explore how new technologies and automation are impacting road safety. The use of these technologies not only enhances safety, but also improves efficiency and reduces environmental impacts. Sweden uses new and unique technologies to improve transport systems. In addition to automated vehicles, traffic surveillance and traffic management are becoming increasingly automated as well. An example of automated surveillance is road safety cameras (ATCs, Automatic Road Safety Control). There is also an increase in automation systems for parking, loading, and unloading. Using a geo-fence, for example, allows authorized vehicles only access to a geographically defined area. It could also restrict vehicles' speeds in a specific area based on the permitted speeds. The authors in [10] try to consider the practical side of implementing new technologies in transport by investigating past good practices and examples of specific implementations. The study [10] explores relationship between the implementation and automation of new technologies and increasing security by reducing the number of human deaths.

The systematic growth of traffic accidents involving pedestrians has been precipitated by factors such as rapid urbanization, an increase in the number of vehicles, speeding, and negligence of road safety [11]. As mentioned in 'Global Status Report on Road Safety' released by World Health Organization [1], traffic accidents are one of the leading causes of death worldwide, with approximately 1.35 million deaths due to road traffic collisions each year. According to the report, pedestrians are the most vulnerable road users, as they account for a large number of road deaths. Due to this increasing importance, the forensic investigation of vehicle-pedestrian accidents is necessary for the detection, investigation, and reduction of road casualties, and it is urgent that this discipline develops permanently. Authors in [11] explore the capabilities and effectiveness of forensic examination in addressing road fatalities, as well as issues that are frequently addressed during such examinations and typical physical evidence used in collision reconstructions between vehicles and pedestrians.

Authors in [12] review different research papers which have analysed various risk factors related to safety of motorized two-wheelers (MTWs). Various countries around the world are seeing a rise in the ownership of MTWs. Particularly high growth rates have been observed in developing countries with high populations and typical traffic characteristics. Consequently, there have been more accidents and deaths. Due to this sudden increase in demand, researchers in [12] decided to conduct studies specifically on MTWs, which behave very differently from cars in terms of their physical and dynamic characteristics. Researchers are further challenged by the unique traffic patterns found in developing countries, since the traditional focus of transportation safety research was based on homogeneous car traffic patterns.

Authors in [13] explore the associations between deprivation and road traffic injury risk for different road user groups. They also review on ways of reducing inequalities in injury risk. Finally, the report surveys current policy and practice across London on addressing deprivation in road safety. As discussed in the report, deaths from injuries are increasingly stratified by social class. For children in the lowest socioeconomic class, the injury-related death rate is five times higher than for those in the highest socioeconomic class. The later updates confirm that these inequalities in road injury risk persist, and even may increase. Children of parents in the lowest socioeconomic classes died at a rate 20 times higher than children of the highest socioeconomic class as pedestrian, a rate 27 times higher as pedal cyclists and a rate 5 times higher as car occupants compared to children of parents in the highest socioeconomic class.

Authors in [14] propose a system which uses electrocardiogram signals and steering wheel motion to monitor drivers' fatigue and drowsiness. The collected data are transferred with a Bluetooth low energy profile. They apply different machine learning (ML) models to detect the patterns of fatigue and drowsiness. As discussed in the paper, the support vector machines classifier could achieve the highest accuracy on the problem. The low-cost proposed prototype gives warnings to drivers regarding their physiological and physical situations which eventually can increase road safety.

As discussed in [15], the average annual daily traffic (AADT) is commonly used as a measure of exposure in regression models of expected crash frequency for road segments and intersections. State and local agencies typically determine ADDT values using short-term traffic counts and extrapolating them over time and space. Estimating traffic volume with such uncertainty can have serious consequences, including: 1) biased estimates of the regression coefficients; and 2) increased dispersion. Researchers and practitioners have difficulty understanding the structure and magnitude of measurement error in AADT estimates, thus making it difficult to account for this error in statistical road safety models and to resolve its problems. This study employs regression calibration and simulation extrapolation techniques to explore how measurement error impacts statistical road safety models by employing measurement error correction approaches like measurement error correction. The authors demonstrate the concept using crash data, traffic data, and roadway data obtained from two-lane vertical curves in rural Washington.

Based on the results, when measurement error correction is applied to regression models predicting expected crash frequency, the regression coefficient estimates with a positive coefficient are bigger and those with a negative coefficient are smaller.

TABLE I
SUMMARY OF LITERATURE REVIEW

Study Title	Author	Contribution
Evaluating the speed camera sites selection criteria in the UK.	Li et al. 2021 [9]	Cameras at different sites are evaluated regarding previous accident history and risk value.
The forensic investigation of vehicle-pedestrian collisions: a review.	Nogayeva et al. 2020 [11]	A discipline review of forensic analysis of vehicle collisions and pedestrian collisions.
Safety of motorized two wheelers in mixed traffic conditions: literature review of risk factors.	Damani and Vedagiri 2021 [12]	Factors affecting safety of two wheelers motorized revealed.
Deprivation and Road Safety in London A report to the London Road Safety Unit	Edwards et al. 2008 [13]	It focused on various relationship and risks related to road safety.
Increasing Road Safety with Machine Learning - A Fatigue and Drowsiness Detection System	Cerca et al. 2020 [14]	Various ML techniques are developed to improve road safety system.
Applications of Measurement Error Correction Approaches in Statistical Road Safety Modelling	Musunuru and Porter 2019 [15]	How measurement error impacts statistical road safety models by employing measurement error correction approaches like measurement error correction
Identification of factors affecting accidents on the intercity road network	Bakalnova et al. 2021 [16]	Examines the effect of different factors that influence the rate of accidents on the intercity road network

Authors in [16] examine the effect of different factors that influence the rate of accidents on the intercity road network. To carry out the research, ten road sections are selected that varied in their level of accidents, traffic characteristics, and geometric parameters. A road experimental study using specialized equipment shows that the accident rates on these roads with similar parameters having been following all standards are strikingly different. Measurements of standard indicators in road sections and accident rates on them do not correlate, indicating that unaccounted factors have an impact on the latter indicator, which is not considered by approved methods.

III. Problem Statement

Nowadays, road safety is a great concerned issue in every country. Traffic accidents have great economic and emotional liability on society. It is obvious that road deaths are not inevitable, but the severity of injury can be reduced by demonstrating effective and comprehensive road safety strategies. Although Great Britain has one of the best road safety records in the world, reports [2], [3] say that every day five people are dying because of road crashes. This study investigates main factors in road accidents and injury severity in Great Britain using STATS19 data expanded by OSM and Visual Crossing. The identified factors help to implement road safety measures for improving road safety strategy.

A. Data Collection

The source of data used in this study is based on STATS19 dataset published by UK government [5]. SATAS19 consists of three different types of datasets including accident, causality, and vehicle, collected in Great Britain from January 2009 until December 2019. They have a common attribute called "Accident Index" for merging them together. The exploratory analysis in this paper is based on the accident dataset with 1,546,493 records. Since this dataset does not include all information for a comprehensive analysis, we expand the location and weather features using OSM and Visual Crossing, respectively. Using OSM and Overpass API, we build a map server and then retrieve all road segments and road nodes and dump them to a database of road data. Then, we clean the road data and expand the map-based static features. The components of the input data that do not vary over time are referred to as static features. There are elements derived from road geometry, such as geohash, segment curvature, orientation, proximity to nearby intersections, as well as number of amenities and shops around a node [3]. In addition, we use Visual Crossing weather API to incorporate weather data into the dataset.

B. Data Pre-processing

As a part of data pre-processing, we remove features with high similarity. To consider the relationship features, we calculate correlation for each pair of features using the equation:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

Possible values of the correlation coefficient range from -1 to +1. Fig. 1 shows a snapshot of feature correlations before cleaning.

As it can be seen, features' correlations can be positive or negative. A negative correlation is a relationship between two features such that as the value of one feature increases, the other decreases. A positive correlation is a relationship between two features that tend to move in the same direction which means when one feature tends to decrease as the other feature decreases, or one feature tends to increase when the other increases. In this respect, a correlation of +1 indicates a perfectly positive linear correlation between two features which means those features are same and one of them can be removed from the dataset. Fig. 2 shows a snapshot of feature correlations after cleaning. For instance, as shown in Fig. 1, the correlation between *Location_Easting_OSGR* and Longitude has the value of +1 and we can remove one of these features. We remove *Location_Easting_OSGR* from the dataset. Similarly, we remove the similar features until there is no two different features with correlation value of +1.

	Location_Easting_OSGR	Location_Northing_OSGR	Longitude	Latitude	Police_Force
Location_Easting_OSGR	1.0	-0.43	1.0	-0.43	-0.37
Location_Northing_OSGR	-0.43	1.0	-0.44	1.0	0.18
Longitude	1.0	-0.44	1.0	-0.44	-0.39
Latitude	-0.43	1.0	-0.44	1.0	0.18
Police_Force	-0.37	0.18	-0.39	0.18	1.0
Accident_Severity	0.03	-0.035	0.03	-0.035	-0.046
Number_of_Vehicles	0.015	-0.033	0.016	-0.033	-0.0022
Number_of_Casualties	-0.04	0.029	-0.039	0.029	0.018
Day_of_Week	0.0002	0.0037	0.00015	0.0037	-0.0017
Local_Authority_(District)	-0.4	0.14	-0.41	0.14	0.98

Fig. 1 Feature Correlation before Cleaning

	Longitude	Latitude	Accident_Severity	Number_of_Vehicles	Number_of_Casualties
Longitude	1.0	-0.44	0.03	0.016	-0.039
Latitude	-0.44	1.0	-0.035	-0.033	0.029
Accident_Severity	0.03	-0.035	1.0	0.073	-0.068
Number_of_Vehicles	0.016	-0.033	0.073	1.0	0.24
Number_of_Casualties	-0.039	0.029	-0.068	0.24	1.0
Day_of_Week	0.00015	0.0037	0.0039	0.0009	0.00029
Local_Authority_(District)	-0.41	0.14	-0.049	0.0023	0.026

Fig. 2 Feature Correlation after Cleaning

IV. Exploratory Analysis

This section represents an exploratory analysis of traffic accidents based on the historical dataset.

The distribution of accident severity is illustrated in Fig. 3. It shows that fatalities are caused by only 1.2% of the total number of accidents, approximately 18,000 out of 1.5 million. Moreover, 15.2% of accidents are considered serious, while 83.6% are classified as light.

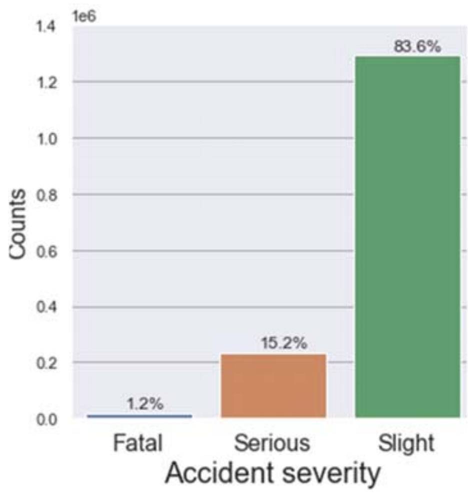


Fig. 3 Severity Distribution

Fig. 4 shows that there are around 165,000 accidents in 2009 and it has been decreased year by year until 2019 which is under 120,000. Additionally, it represents the reduction rates for each year rather than the previous year e.g., there is 4.2% reduction in accidents in 2019 compared to 2018. In 2014, there is a sudden increase in accidents; however, they decreased again in the following years.

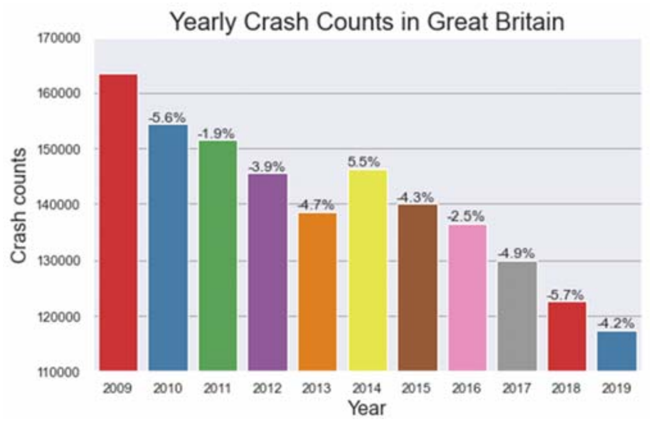


Fig. 4 Yearly Crash Counts

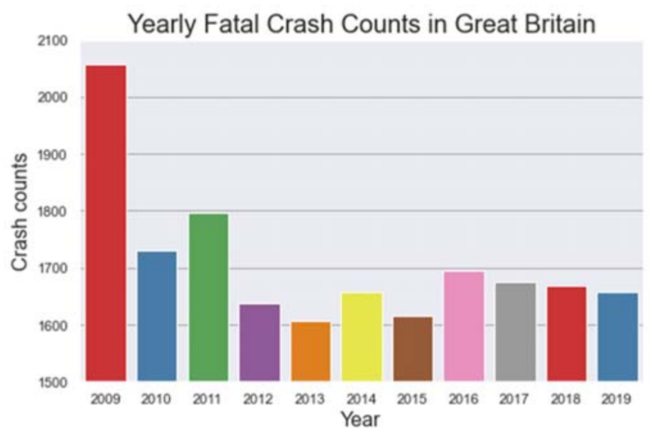


Fig. 5 Yearly Fatal Crash Counts

Fig. 5 shows the number of accidents caused at least one death. As you can see, the number of fatal is decreasing over the years but there is no similar trend as Fig. 4. This means there is no same trend downswing in yearly fatal distribution shown in Fig. 5.

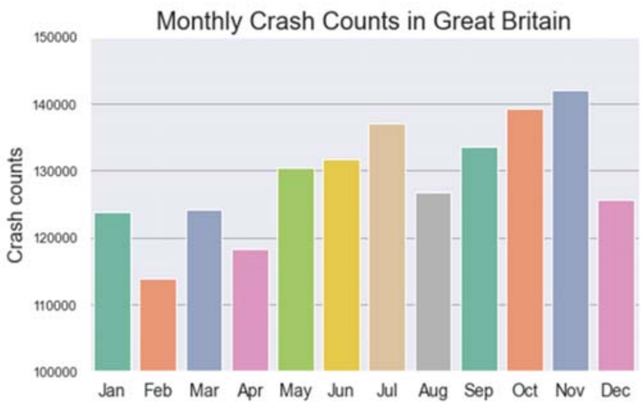


Fig. 6 Monthly Crash Count

Fig. 6 shows that there are around 145,000 accidents in November, and it has been decreased by December which is 125,000. As it can be seen, there is a greater number of crashes during the months of July, September, October and November which are more than 130,000. There is a sudden decrease in February and April which does not long last. The reasons can be explained by different factors like the number of people and vehicles on roads due to holidays and climate changes. Additionally, weather conditions can cause road accidents as well.

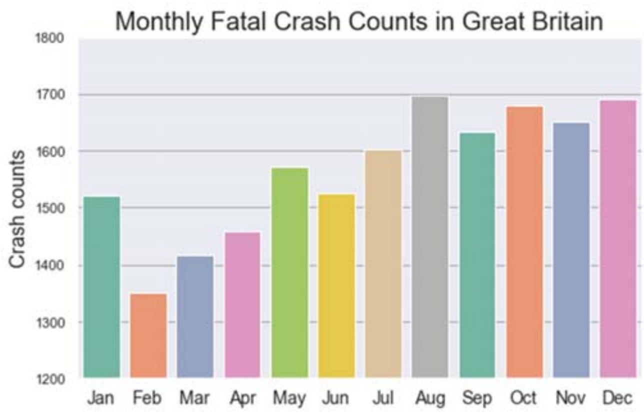


Fig. 7 Monthly Fatal Crash Count

Fig. 7 shows the number of fatal crash counts for each month. As you can see, the number of fatal is gradually increasing over the months but there is no similar trend as Fig. 6. There are greater numbers of fatal crashes between July and December.

The Monthly Fatal Crash Count shows an increasing graph over cold, foggy and snowy months when failing to follow traffic laws causes higher number of fatal crashes.

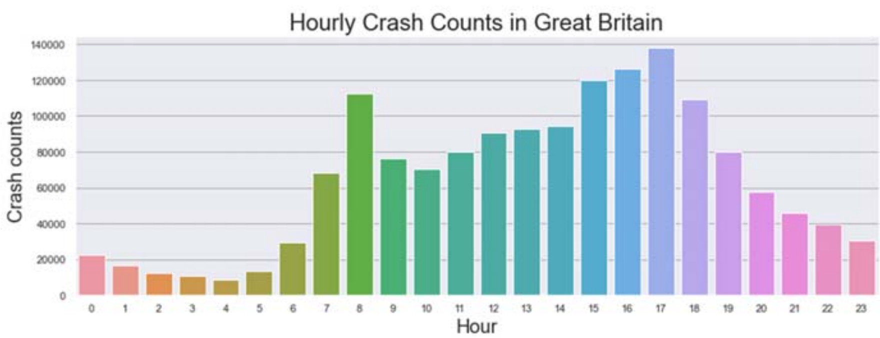


Fig. 8 Hourly Crash Count

Fig. 9 Hourly Fatal Crash Count

Fig. 8 shows the hourly crash counts that are above 100,000 during peak hours in the morning 8:00 and in the evening from 15:00 till 18:00. The number of crashes gradually decreases at each hour after 18:00 until the next day morning at 7:00.

The hourly crash counts show that most accidents are in the rush hours which are one hour in morning and about three hours in the evening where traffic congestion on roads is at its highest because too many people are on the road to go/come back to/from work.

Although the UK has the most clean and safe roads with lots of signals and speed limits, there is still a considerable number of fatal crashes specially in rush hours. Fig. 9 shows the hourly fatal crash counts over 24 hours. As it can be seen, there is a high range of fatal crashes from 10:00 to 17:00 increasing timely and then gradual decrease from 18:00 to 9:00.

Fig. 10 shows hourly crash counts that are high from Monday to Friday and then gradually decreases over the weekends. The hourly crash counts are higher at evenings on working days when many people come back from work and on Friday when there are lots of drunk people on roads which may not follow the guidelines properly.

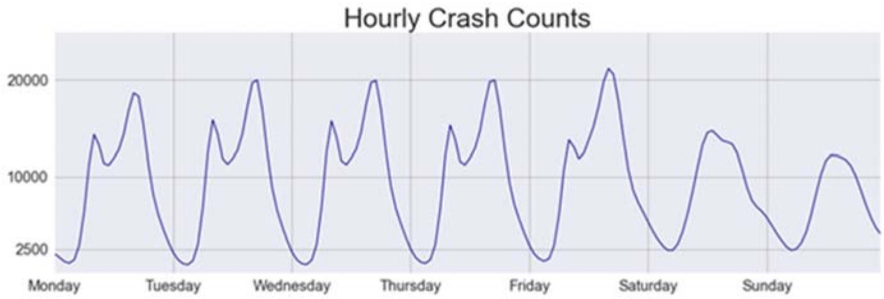


Fig. 10 Hourly Crash Count during Week

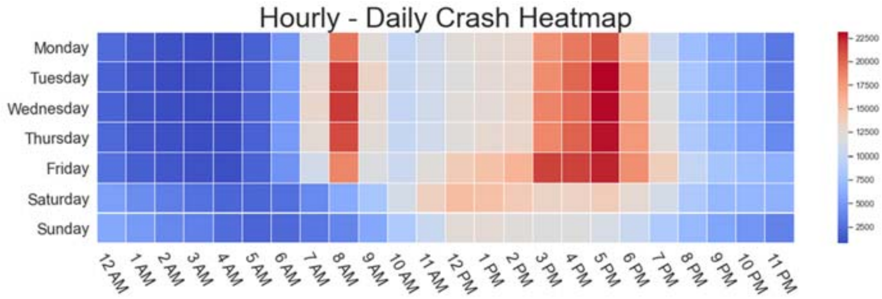


Fig. 11 Daily Crash Heatmap

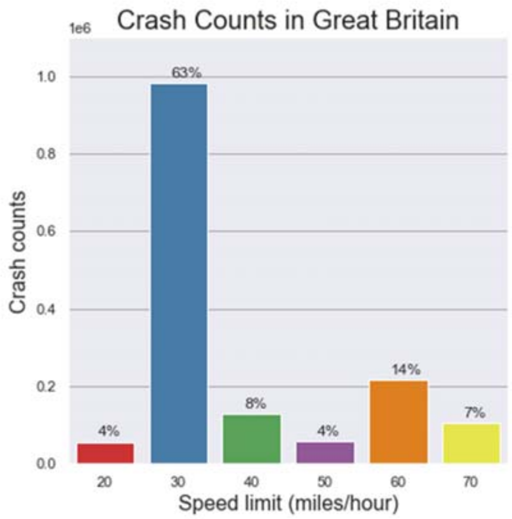


Fig. 12 Crash Count in Great Britain

Fig. 11 shows hourly-daily crash heatmap. As it can be seen, there are more crashes during rush hours: 8:00 in the morning and 15:00-18:00 in the evening. The crash congestion reduces hourly after 19:00 until the next day morning around 7:00. Then, it jumps in its highest at 8:00 and 17:00 and also during Fridays.

The hourly-daily crash heatmap shows that most accidents are in the morning one hour and evening about three hours when there is high road congestion because more people travel between work and home. Out-of-the-ordinary road congestions can cause more accidents as there would be more likely to get people who do not respect the traffic laws.

Fig. 12 shows the crash counts and crash rates versus speed limits. As it can be seen, 63% of crashes are related to the roads with the speed limit of 30 mph and 14% with the speed of 60 mph. The crash count is at its highest in 30 mph speed limit because there are more urban roads with this speed limit which are also very tight and narrow somewhere in the UK, specially in London. Due to traffic congestions in large cities, there are more likely for people to mistake during driving for example on crossways by not giving proper signals.

Fig. 13 shows the fatal crash counts and crash rates versus speed limits. As can be seen, 33.6% of crashes are related to the roads with the speed limit of 30 mph and 35.6% with the speed of 60 mph. There are greater numbers of fatal crashes at speed limit of 30 mph and 60 mph. The reason can be explained separately as below:

- 30 mph: Since there are more crashes on roads with speed limits of 30 mph a shown in Fig. 12, based on statistics probability, there must be more likelihood to have more fatal crashes.

- 60 mph: As shown in Fig. 12, the number of crashes is considerably less than the number of crashes with speed limit of 30 mph. However, there is more likely to have fatal crashes because of having higher speed limit.

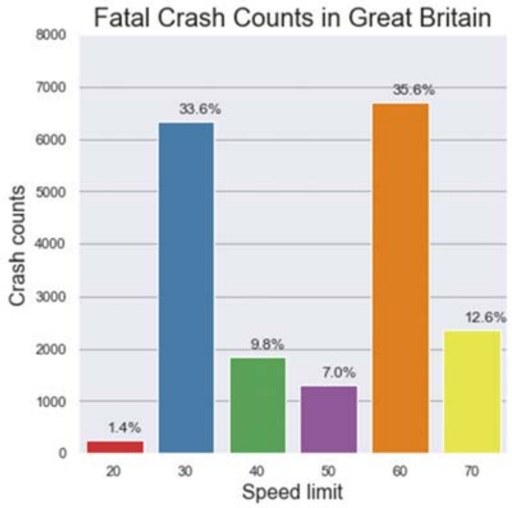
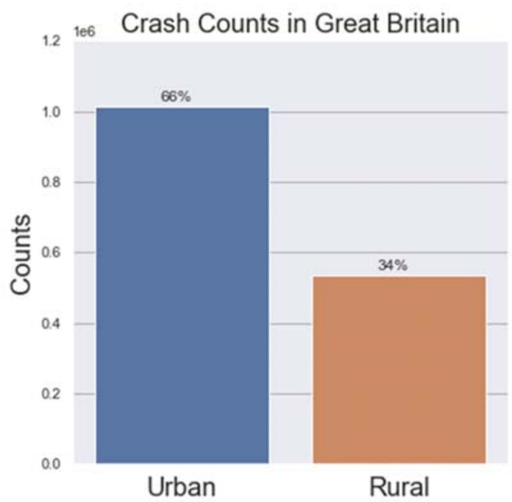
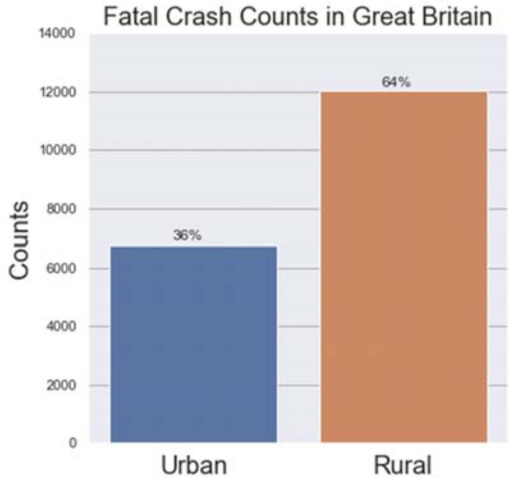


Fig. 13 Fatal Crash Counts vs Speed Limits

Figs. 14 (a) and (b) show crash counts and fatal crash counts for urban and rural areas, respectively. As it can be seen, the number of crashes in the urban area with the crash rate of 66% is more than rural area with crash rate of 34%. Fig. 14 (b) shows the opposite results for fatal crashes which means the number of fatal crashes in rural area with the rate of 64% is more than the number of fatal crashes in urban area with rate of 36%. This result can be explained as: There is less traffic in rural areas and people usually drive at the maximum speed or even higher. Therefore, it would be likely to have more fatal crashes. In contrast, there is usually more road congestion in urban area which gets people to drive slower which results in less fatal crashes. However, they make more chance to make mistakes in narrow roads, which causes a greater number of crashes.



(a)



(b)

Fig. 14 Fatal Crash Count for Urban and Rural Areas

Fig. 15 shows the accident severity versus the weather conditions in Great Britain. The result is very interesting as 80.5% of crashes happen in fine weather when there is no high wind. The next highest rate of crashes which is 11.6% happens in raining weather with no winds. As it can be seen in the figure, there are much smaller number of crashes in other climatic conditions when it is snowing or foggy because drivers are very careful.

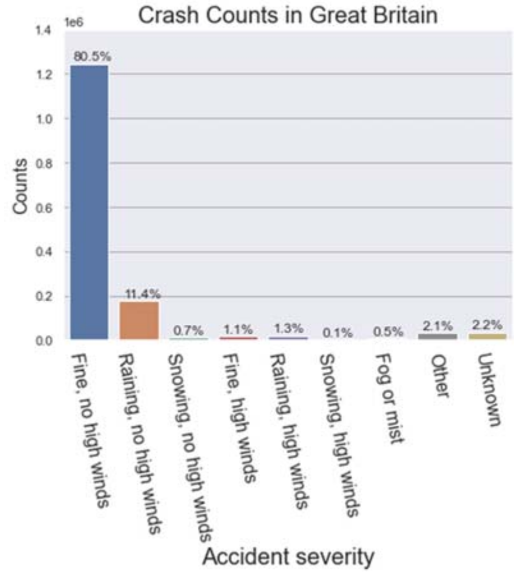


Fig. 15 Crash Counts in Great Britain

Fig. 16 shows the fatal crash count in different weather conditions. As can be seen, 82.6% of fatal crashes happen in fine weather with no high winds. Fatal crashes drop to 9.8% in rainy weather with no wind. There is no fatal crash in snowing weather with high winds because people drive carefully and very slowly. In addition, people may only drive a little compared to acceptable weather conditions. Therefore, a smaller number of cars means a smaller number of crashes. As shown in Fig. 16, fine weather is one of the major causes of fatal accidents because most people use their vehicles to commute instead of public transport, which results in high traffic on the road. In addition, more people ignore safe driving by not following the traffic rules or driving at maximum speed.

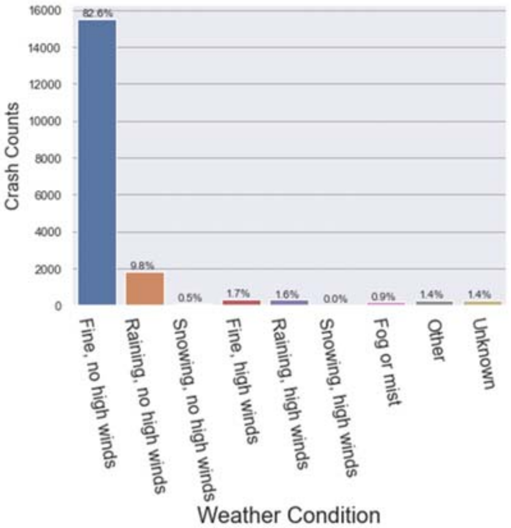


Fig. 16 Fatal Crash Count in Great Britain

Fig. 17 shows the crash count in different light conditions across Great Britain. As it can be seen, the highest rate of crashes, which is 73%, happens in the daylight time and the least crash rate which is 0.6% observed in darkness with lights unlit. In addition, the crash count is slighter high to 19.7% when there is darkness with lights lit.

Crash counts is at its highest rate in daylight because of having more traffic on the roads and people are used to drive at higher speed in daytime.

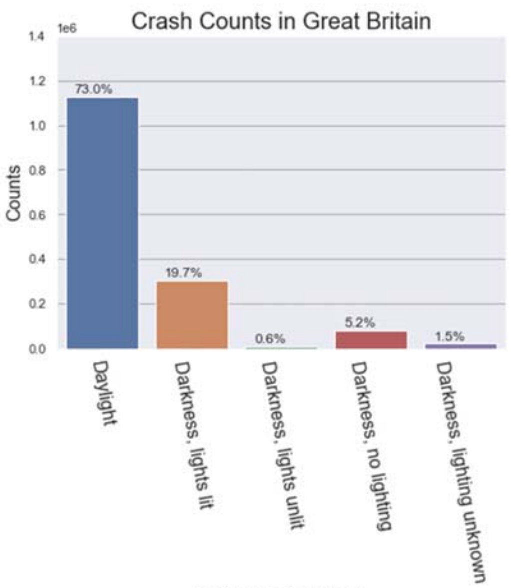


Fig. 17 Crash Counts in Great Britain

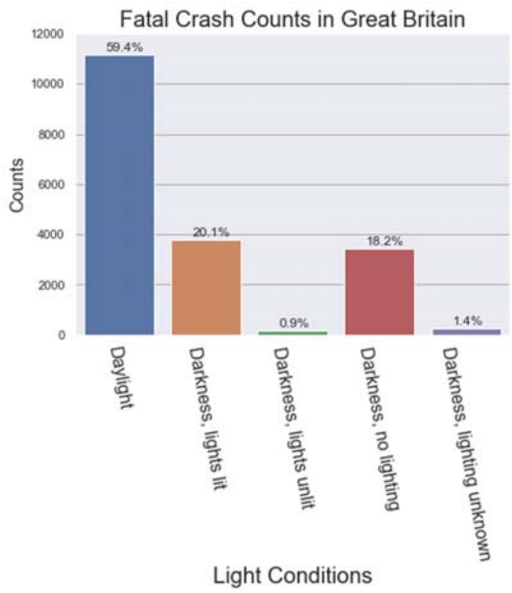


Fig. 18 Fatal Crash Counts vs Light Conditions

Fig. 18 shows the fatal crash counts in different light conditions. As it can be seen, 59.4% of fatal crash counts happen in daylight and the least fatal crashes which is 0.9% happen in darkness when lights are unlit. The results show similar rates of fatal crashes in darkness when lights are lit and in darkness with no lightning, as 20.1% and 18.2% respectively.

The fatal crash counts in different light conditions can be analysed with similar reasons explained for Fig 17. Fig. 18 shows more fatal accidents in daylight time because of the high traffic on roads. In other two conditions when there is darkness with lights lit or with no lightning, the fatal counts are very similar because there is no clear view of roads and people might be drunk or they feel sleepy and fatigue in the night time.

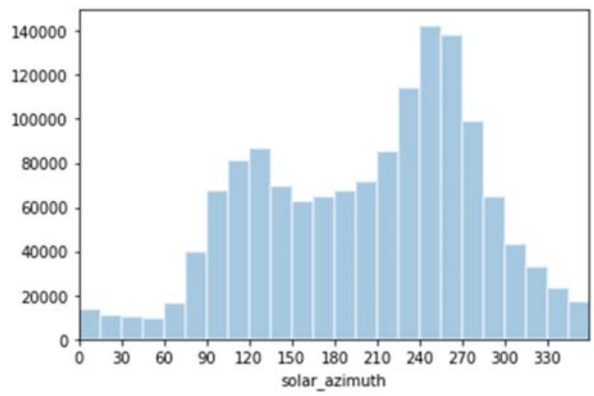


Fig. 19 Crash Counts vs. Solar Azimuth

Fig. 19 shows the number of crashes versus solar azimuth (angle of the sun position). As it can be seen, the crash count is much higher when the solar azimuth is between 240 to 270. The least crash count is observed when sun position is between 0 to 70 and 300 to 330. Crash counts are still large between 60000 to 80000 when the solar azimuth is between 90 to 300.

Solar azimuth produces different reflection of sun lights on roads at different time. It changes with time and location. If drivers face sun light opposite to driving direction, their vision can be impaired and driving performance can be affected.

Fig. 20 shows crash counts versus solar altitudes. As it can be seen, more than 13000 crashes happen when the solar altitude is 10. The number of crashes gradually increase from solar attitude of -60 to 10 and then decreases from 10 to 60. The least crash count is observed when the altitude is at -60.

Direct light reflection from sun on the earth will be based on different solar azimuth and attitude.

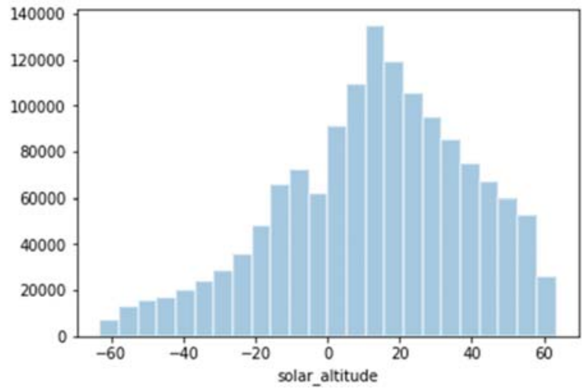


Fig. 20 Crash Counts vs. Solar altitude

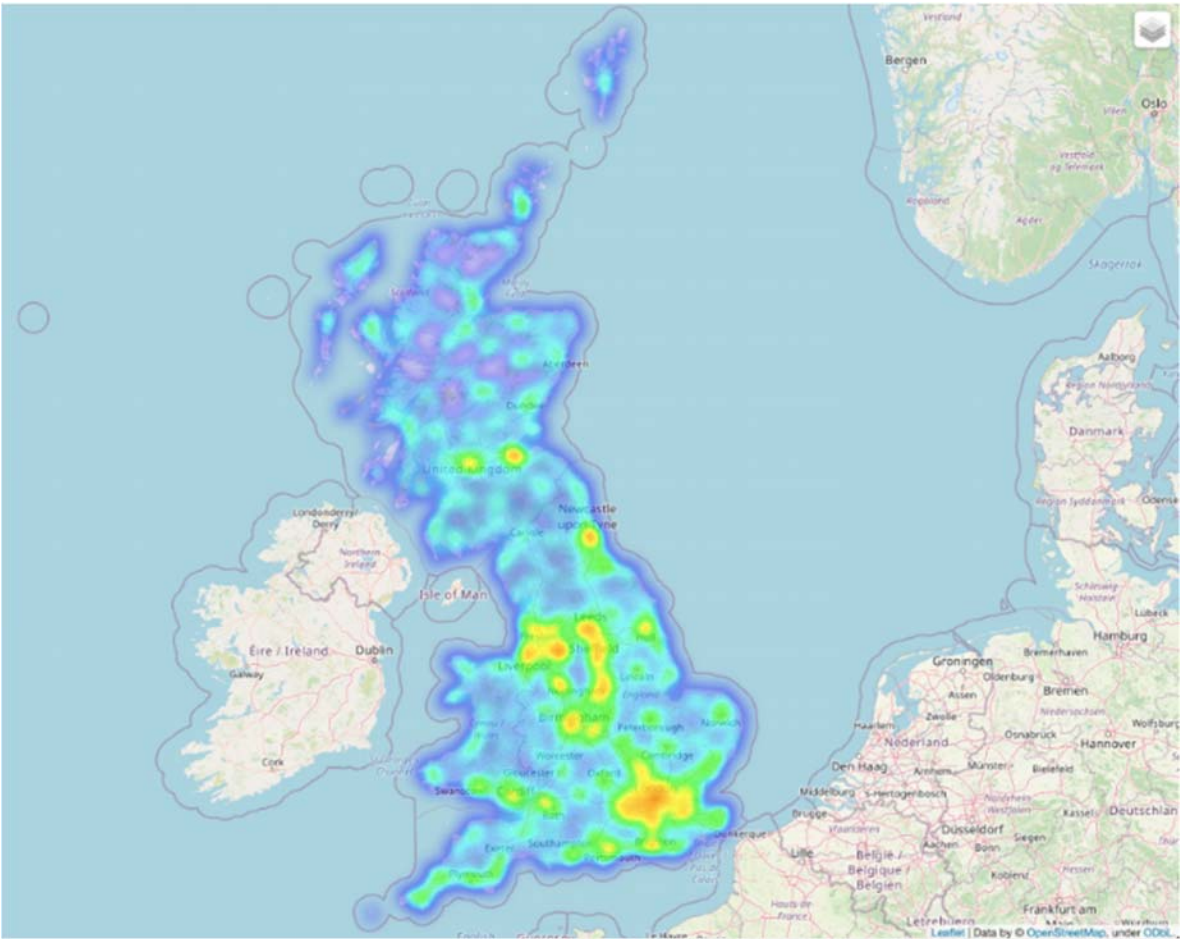


Fig. 21 Crash Counts Heatmap

Fig. 21 shows the crash counts heatmap across the Great Britain. As it can be seen, there are more crashes in London and area nearby to Leeds, Sheffield, Liverpool, Nottingham, Brighton because of more density of vehicles across these regions. The smaller numbers of crashes are observed in Scotland, Wales, Plymouth, Cardiff, and Exeter. Few areas in central London are more prone to crashes than others. For example, traffic accident occurs almost every six weeks around Camden Town, Elephant, and Castle, especially nearby underground stations. In addition, a fatality results every year in one of these areas. This indicates that road crashes may be triggered by poor road conditions, poor lighting or poorly laid out roads [17].

Fig. 22 shows the number of road crashes in Great Britain counties. As it can be seen, around 100,000 crashes are reported in Brighton and Surrey. In Norfolk and Suffolk counties, 80,000 crashes are reported which is also a high number compared to other areas like Plymouth, Scotland, and Wales. Traffic volume is increasing primarily because the economy is growing. As traffic volume increases, incidents also increase. As such, accidents continue to increase until economic development reaches a critical threshold. In order to counteract the effect of increased traffic, it becomes imperative to invest in better training, better vehicle standards, enforcement, and engineering. In this manner, even if traffic volumes continue to grow, the number of incidents and deaths begins to decline.

Fig. 23 shows the number of road crashes around London. The number of crashes is about 100,000 in Oxted, Dorking, Guildford, Crawley, Redhill, East Grinstead, Woking, Epsom, and Stratford which is surpassingly high. Moreover, less crash count, between 0 to 10,000 is observed in areas of Wimbledon, Croydon, Begging Hill and Sutton. Remaining areas lie down between 20,000 to 40,000 crash counts.

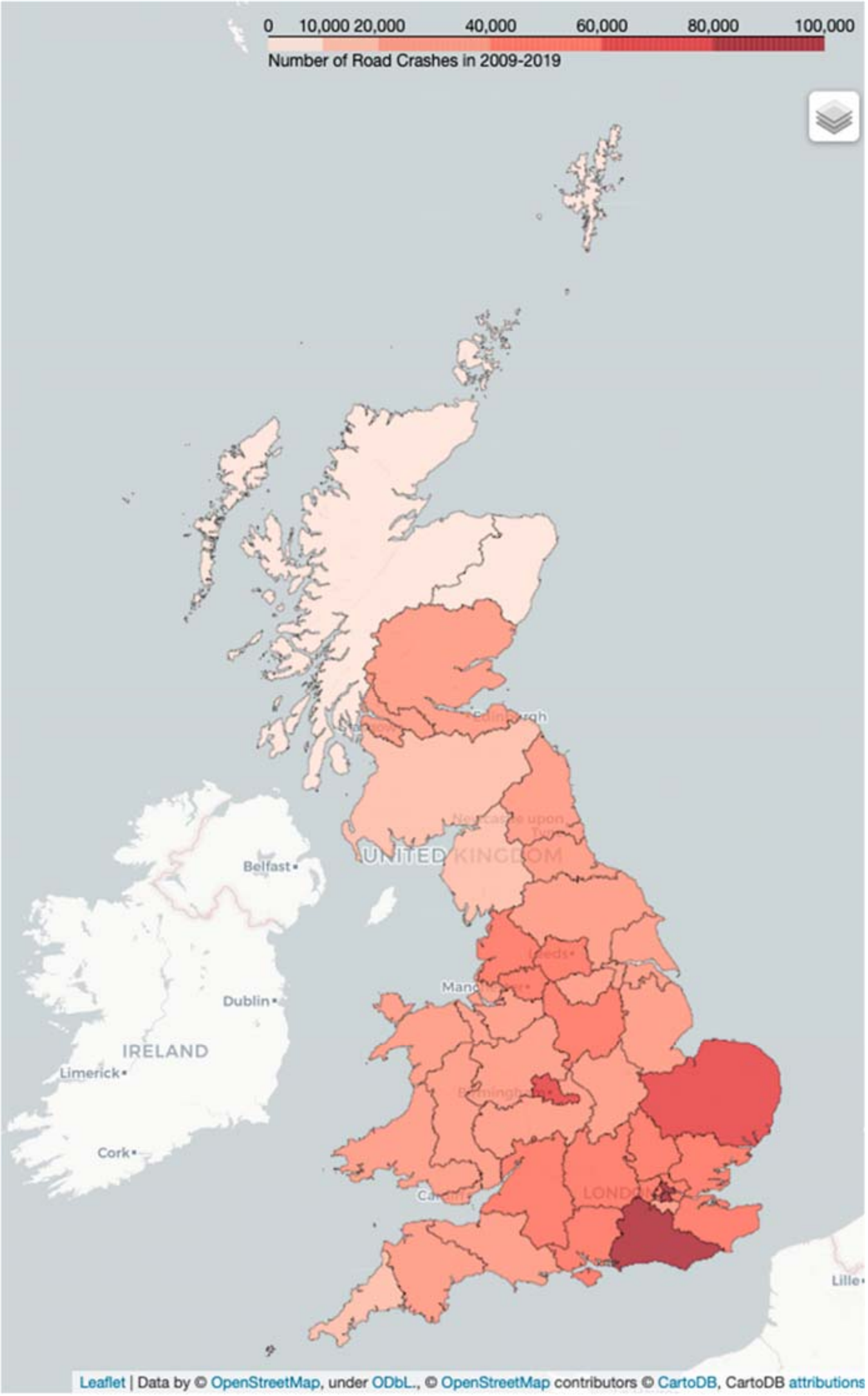


Fig. 22 Number of Road Crashes in 2009-2019

A total number of 8,534 fatal traffic accidents are recorded across some of the busiest roads in Great Britain. The A6 between Luton and Cumbria topped the list of Britain's most dangerous roads, with 70 fatalities. The A5, connecting London to Holyhead is found to be the second one, with 67 fatal accidents and the A40 from London to Fishguard is close behind with 65 fatalities.

Fig. 24 shows the number of fatal crashes in Great Britain counties. As it can be seen, the fatal crash count is high with the value of 1,200 in Norfolk and Suffolk counties and then 800 in Brighton and Surrey. The fatal count is between 0 to 200 for some areas like Plymouth, Scotland, and Wales.

Rutland in England's East Midlands has the nation's highest rate of fatal traffic accidents, with 5.181 per 10,000 residents, followed by Powys in Wales (4.312), Fermanagh & Omagh in Northern Ireland (4.150) and the Orkney Islands in Scotland (4.110). Clackmannanshire, Scotland's smallest county, is the safest area with no fatalities between 2012 and 2016. Cardiff, Tyne and Wear and Belfast City have the next lowest rates per 10,000 residents.

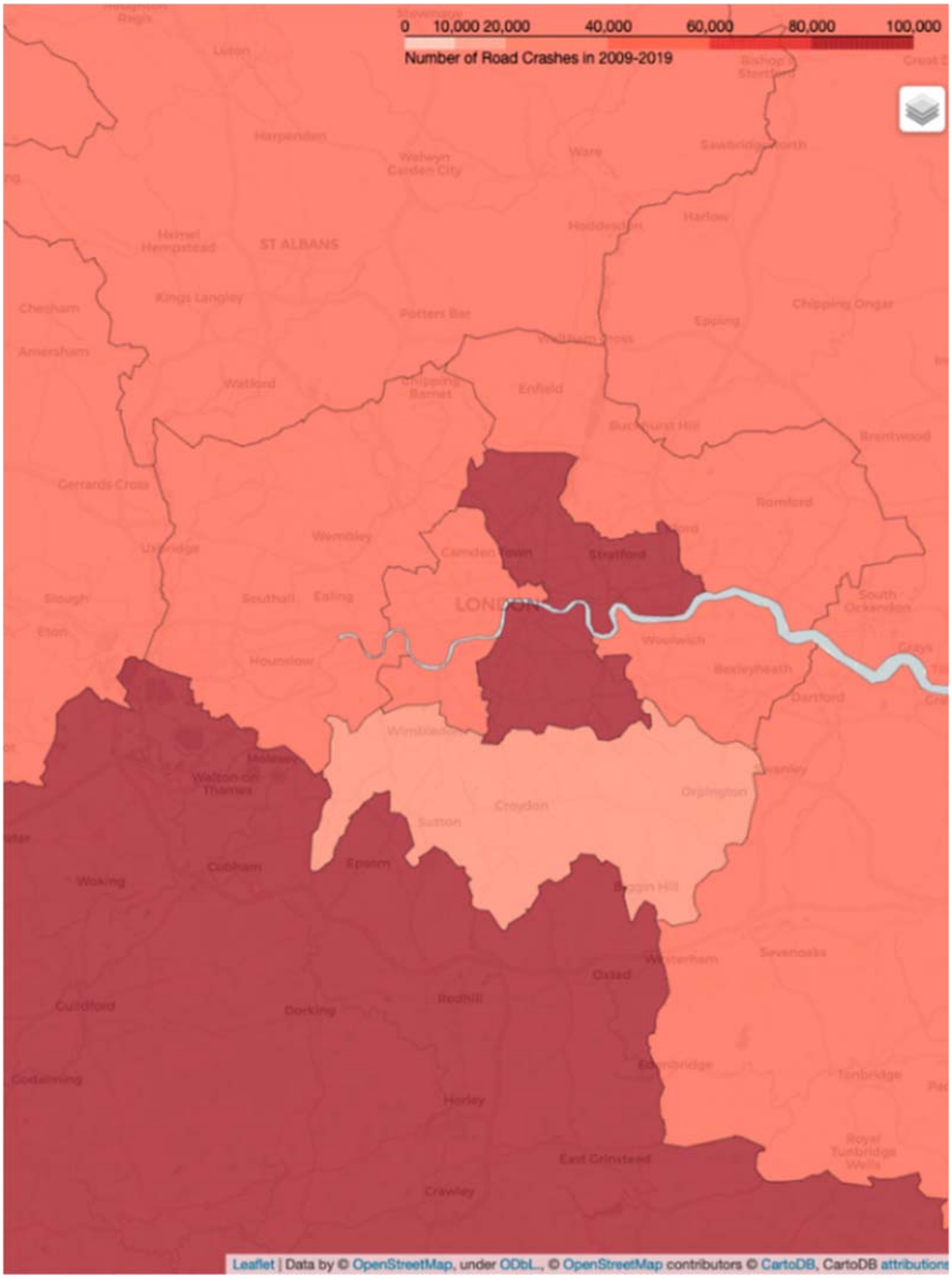


Fig. 23 Number of Road Crashes around London in 2009-2019

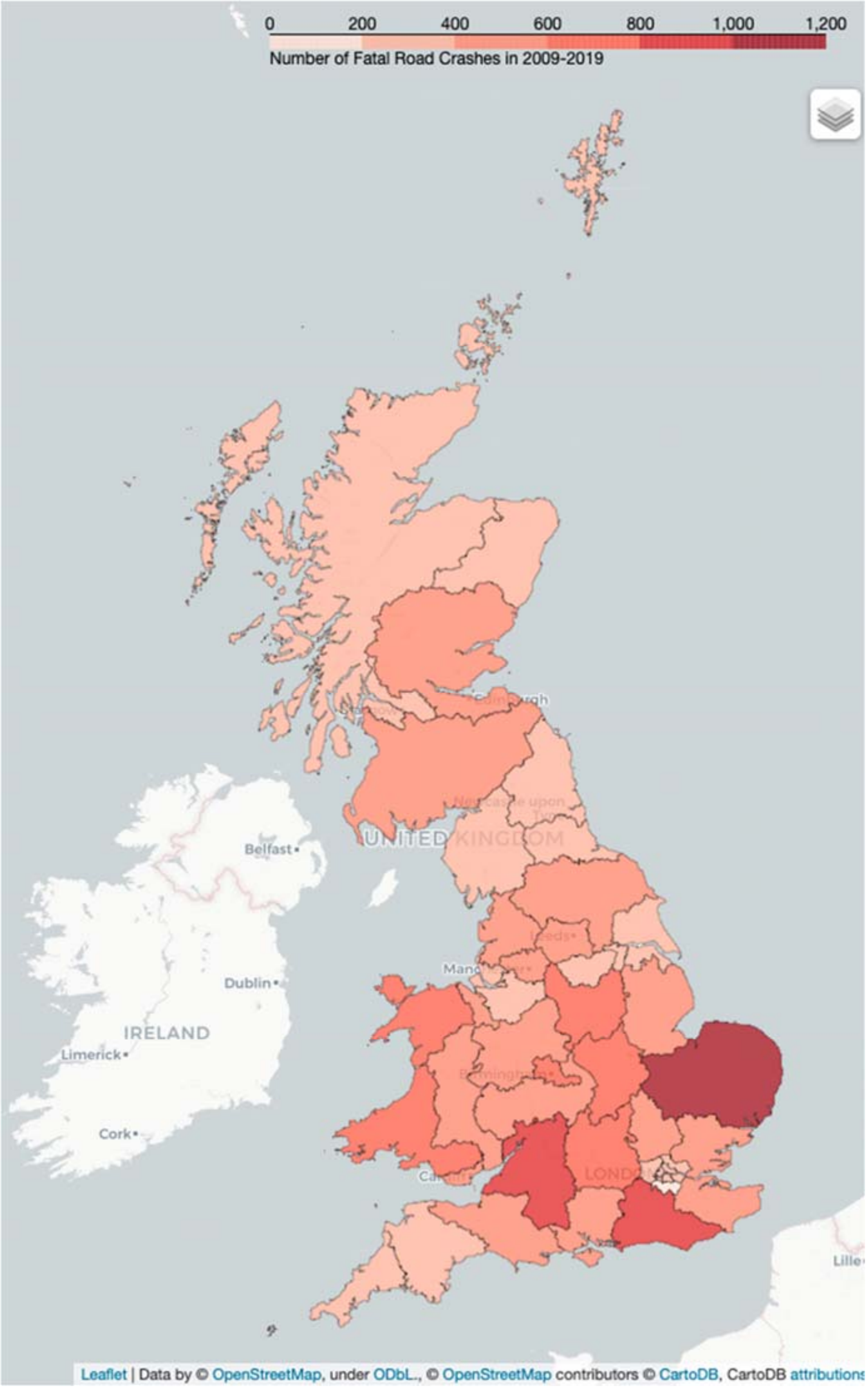


Fig. 24 Number of Fatal Road Crashes in 2009-2019

Fig 25 shows the number of fatal road crashes around London from 2009 to 2019. As it can be seen, the number of crashes is near to 1200 in Oxted, Dorking, Guildford, Crawley, Redhill, East Grinstead, Woking, Epsom, and Stratford which is surpassingly high. Moreover, less fatal counts are observed in area of Wimbledon, Croydon, Begging Hill and Sutton between 0 to 200. Remaining areas lie down between 200 to 600 fatal crash counts.

There are 25,341 reported collisions in London in 2019. People who walk, cycle and motorcycle make up 83% of all people killed or seriously injured in 2009 to 2019.

Fig. 26 shows the top 10 most risky places around London. As it can be seen, the area near to Whitechapel is one of the riskiest places with 33 fatal crashes. There is a smaller number of fatal crashes, about 23, in Woolwich, Streatham, and Walthamstow. The number of pedestrians killed on London's roads increased to 68 in 2019, compared to 57 in 2018, representing 54% of overall fatalities; 44 of these injuries are the result of a collision with a vehicle. In 2019, cycling fatalities decreased from 12 to five; motorcyclist fatalities increased from 22 to 31; and 3,780 people were injured seriously [18].

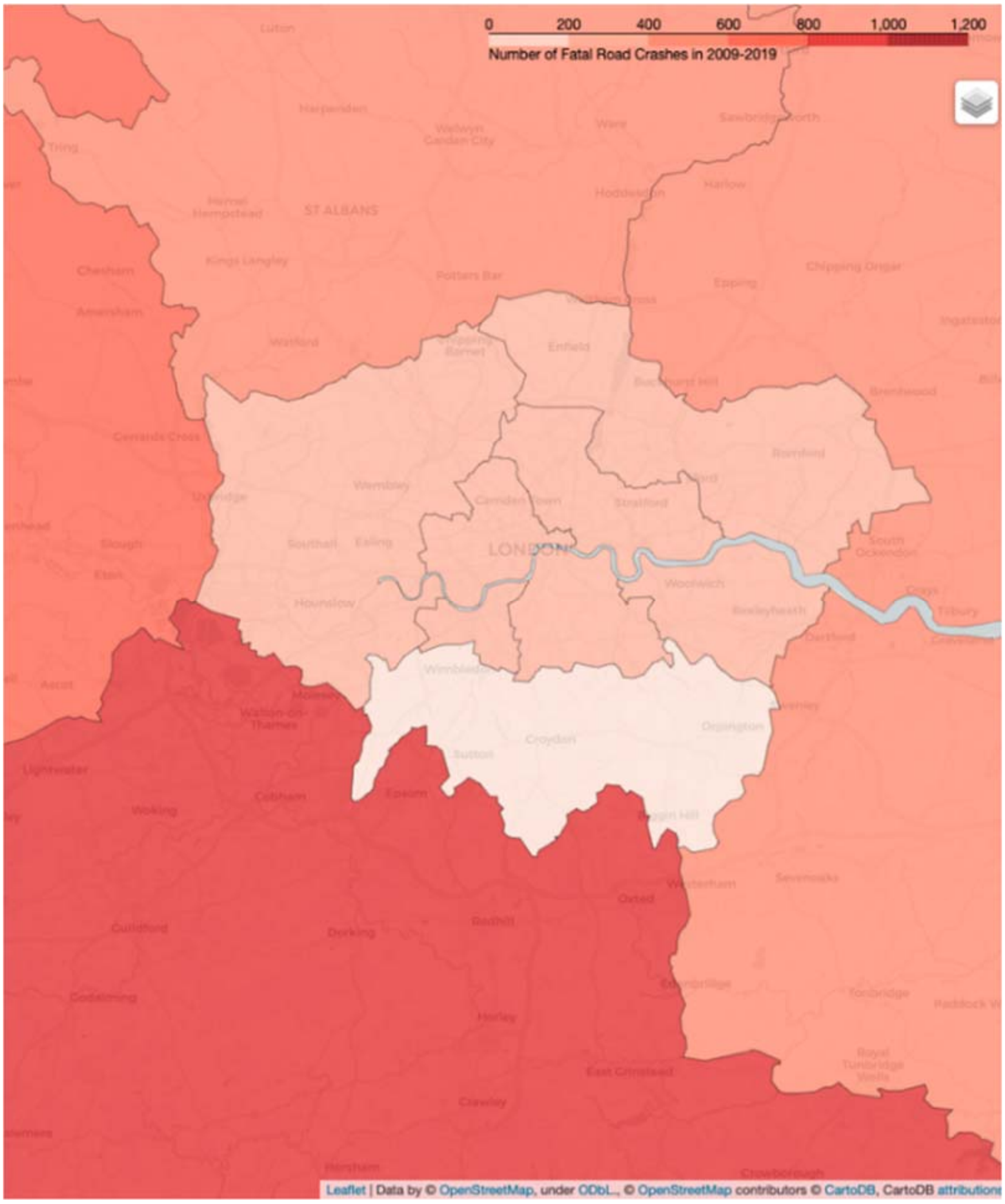


Fig. 25 Number of Fatal Road Crashes around London in 2009-2019

V. Discussion

Accident severity has been explained with the risk percentage, which is categorized into fatal, serious, and slight. This study explored the conditions of accidents and casualties which could provide policy makers with invaluable insights into how accidents happen and how they can be mitigated. For example, the road traffic analysis can help policy makers to:

- construct roads to divert traffic from congested area,
- investigate high density points in map geolocation analysis to discover construction needs,
- increase response time during rush-hours,
- find out the important factors on road safety,
- recognize dangerous junctions which cause specific manoeuvre.

Road fatalities are not caused by a single factor. As reported in [19], many factors influence them such as:

- Travel distances (partially affected by economic externalities).
- Driver, rider, and pedestrian behaviour.
- Use of various modes of transport.
- Effects caused by external factors, such as weather (e.g., encouraging/reducing travel or closing roads) or changes in the road surface (making it more slippery).
- A change in the mix of people using the road (for example,

more newly qualified drivers and older drivers).

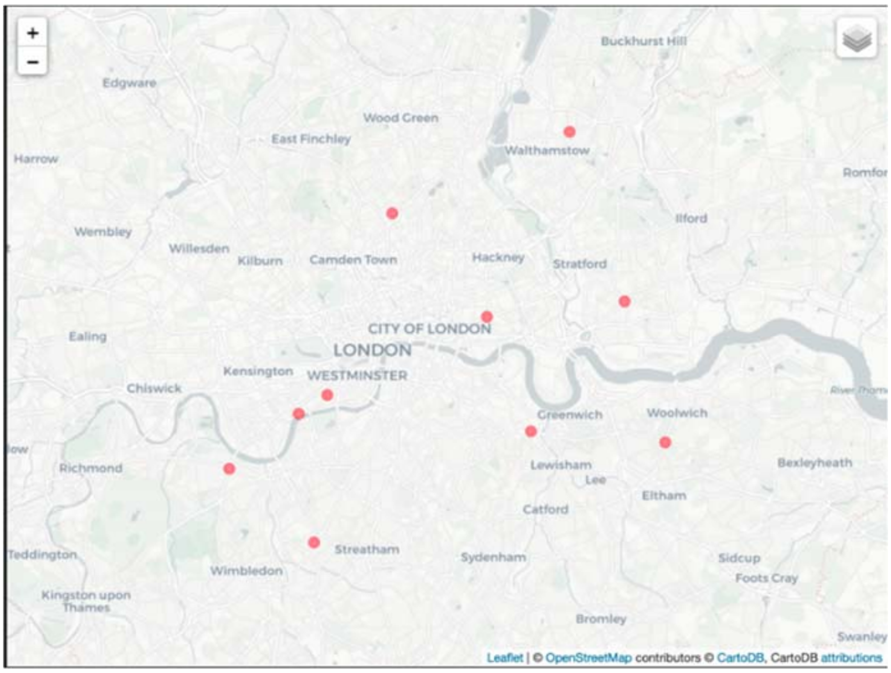


Fig. 26 Top 10 most risky places in London

VI. Possible Solutions

We have outlined the proposed solution here which could be beneficial in reducing the road accidents in the future and to overcome all the risks that result in accidents in various conditions. There are many latest technologies available in the market using ML, AI, IOT etc. which results in smart highways and roads. Innovative technologies for road safety, including Intelligent Transport Systems (ITS), transport telematics, advanced driver assistance technologies, and, more recently, eSafety, reflect the growing use of electronic and telecommunication technology in the road transport industry [20]. A few aspects of advanced technology have the potential to significantly impact road safety, including the development of driving automation and connected transportation systems.

The IoT and ML are two cutting-edge technologies that will certainly become increasingly significant in our daily lives in the coming years. We can save costs, minimize damage, and manage hazards in our transportation industry by incorporating IoT technologies. The use of connected sensors in conjunction with ML-powered analytics tools can assist us in collecting data, analysing it, making predictions, and making decisions that will keep our roads safe [21]-[23]. The major groups of ML techniques identified for road safety modelling are nearest neighbour classification, decision trees, evolutionary algorithms, support-vector machines, and artificial neural networks.

IoT is used to make smart roads safer, more efficient, and environmentally friendly. Smart roads rely on both physical infrastructures such as sensors and solar panels as well as software infrastructure such as artificial intelligence and big data [9].

Embedded technologies in smart roads can monitor road conditions, communicate with autonomous and connected vehicles, generate energy and more. Some examples are:

- IoT connectivity: Using IoT devices, cities can gather data on traffic and weather on the road and reduce traffic congestion, manage traffic, and increase energy efficiency through this style of connectivity.
- Traffic management networks: The network uses speed cameras to provide warnings of hazardous conditions and automates traffic diverting signals to reduce congestion and increase safety.
- Traffic lights optimization: Traffic lights can be optimized, and commuters are informed by data from closed-circuit television (CCTV) cameras or smart vehicles.

Most smart transport technical solutions aim at individual vehicles, but there have been major developments in scaling up smart infrastructure. Future urban transport will be smarter due to worldwide experiments in Vehicle to Infrastructure (V2I), Vehicle to Vehicle (V2V), and Vehicle to Pedestrian (V2P) technologies. Here are some examples of new in-vehicle technologies:

- Electronic stability control,
- Active cruise control with emergency brake,
- Blind spot monitoring,
- Brake assist,
- Adaptive headlights,
- eCall,
- Obstacle and collision warning,
- Advance hazard warning,
- Lane departure warning,
- Seat belt reminders.

VII. Conclusion and Future Work

The idea of this study is to make an exploratory analysis of road crashes in the UK. Since the data published by the UK government do not include all information for a comprehensive analysis, we expand the location and weather features using OSM and Visual Crossing, respectively. Using OSM and Overpass API, we build a map server and then retrieve all road segments and road nodes and dump them to a database of road data. Then, we clean the road data and expand the map-based static features. The components of the input data that do not vary over time are referred to as static features. There are elements derived from road geometry, such as geohash, segment curvature, orientation, proximity to nearby intersections, as well as number of amenities and shops around a node [7]. In addition, we use Visual Crossing weather API to incorporate weather data into the dataset. Then, we represent an exploratory analysis with deep insights which could provide policy makers with invaluable insights into how accidents happen and how they can be mitigated. We have come across different new strategies for road safety. This work discusses the accident eradication and safety measures for heavy traffic areas and analyses decrease/increase rates for both road crashes and fatal crashes across different geospatial and environmental conditions to provide deep insights on accident-prone areas.

Some new technologies such as smart traffic lights, traffic control systems, artificial intelligence, and telematics devices can help prevent and reduce traffic accidents and improve road safety. High-speed data transmission can be used, for example, to collect data on traffic, road and weather conditions to real-time warnings to drivers through online road risk assessment.

References

[1] https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
[2] https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2021/reported-road-casualties-great-britain-annual-report-2021
[3] Reported road casualties Great Britain: 2015 annual report, https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2015
[4] https://www.racfoundation.org/research
[5] Road Safety Data. https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data
[6] https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data/datafile/36f1658e-b709-47e7-9f56-cca7aefeb8fe/preview
[7] https://patents.google.com/patent/WO2021191168A1/en
[8] Silva, P.B., Andrade, M. and Ferreira, S., 2020. Machine learning applied to road safety modeling: a systematic literature review. Journal of traffic and transportation engineering (English edition).
[9] Li, H., Zhu, M., Graham, D.J. and Ren, G., 2021. Evaluating the speed camera sites selection criteria in the UK. Journal of safety research, 76, pp.90-100.
[10] Tonhauser, M. and Ristvej, J., 2021. Implementation of new technologies to improve safety of road transport. Transportation research procedia, 55, pp.1599-1604.
[11] Nogayeva, S., Gooch, J. and Frascione, N., 2020. The forensic investigation of vehicle-pedestrian collisions: a review. Science & Justice.
[12] Jaikishan Damani, Perumal Vedagiri, "Safety of motorised two wheelers in mixed traffic conditions: Literature review of risk factors", Journal of Traffic and Transportation Engineering (English Edition), Volume 8, Issue 1, 2021, Pages 35-56
[13] Edwards, Phil, Judith Green, Ian Roberts, Chris Grundy and Kate Lachowycz. "Deprivation and road safety in London - a report to the London Road Safety Unit." (2008).
[14] Cerca, A.; Ferreira, A.; Lourenço, A. "Increasing Road Safety with Machine Learning - A Fatigue and Drowsiness Detection System", Proc Portuguese Conf. on Pattern Recognition - RecPad, Evora, Portugal, Vol., pp. -, October, 2020.
[15] Musunuru A, Porter RJ. Applications of Measurement Error Correction Approaches in Statistical Road Safety Modeling. Transportation Research Record. 2019;2673(8):125-135. doi:10.1177/0361198119841856
[16] Baklanova, K V; Voevodin, E S; Fomin, E V; Kashura, A S; Cheban, E P. "Identification of factors affecting accidents on the intercity road network", IOP Conference Series. Materials Science and Engineering; Bristol Vol. 1061, ISS. 1, (Feb 2021). DOI:10.1088/1757-899X/1061/1/012005
[17] https://www.ucl.ac.uk/mathematical-physical-sciences/news/2018/aug/half-london-car-crashes-take-place-5-citys-junctions#:~:text=Transport%20for%20London%20figures%20show,where%20a%20road%20accident%20is
[18] https://www.fleetnews.co.uk/news/fleet-industry-news/2020/09/30/tfl-data-shows-12-increase-in-london-road-deaths
[19] Reported road casualties in Great Britain: 2019 annual report, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/922717/reported-road-casualties-annual-report-2019.pdf
[20] https://ec.europa.eu/transport/road_safety/statistics-and-analysis/statistics-and-analysis-archive/esafety/vehicle-technologies-and-road-casualty-reduction_en
[21] M.P. Basgalupp, A.C.P.L.F. Carvalho, R.C. Barros, et al. "Lexicographic multi-objective evolutionary induction of decision trees", International Journal of Bio-Inspired Computation, 1 (1) (2009), pp. 105-117
[22] O. Chapelle, B. Scholkopf and A. Zien, Eds., "Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) (Book reviews)," in IEEE Transactions on Neural Networks, vol. 20, no. 3, pp. 542-542, March 2009, doi: 10.1109/TNN.2009.2015974.
[23] K.M. Decker, S. Focardi, Technology Overview: a Report on Data Mining Technical Report CSCS TR-95-02 Swiss Scientific Computing Center, Bern (1995)