Task 2: Research Project Plan:

Road Traffic Accident Risk Prediction in the UK.

**.**

Rita Chinwenwa Uzoka

30034087

# Introduction and justification

The World Health Organization estimates that 50 million individuals are impacted by and injured in traffic accidents each year, killing 1.2 million people. A multitude of issues, including traffic congestion, air pollution, and traffic accidents, have been brought on by the explosion of vehicles in modern society because of the rapid rise of urbanization. Both significant economic losses and human casualties have been brought on by these issues. Around 1.25 million people die in traffic accidents each year, according to the World Health Organization's Global Status Report on Road Safety, which was published in 2015. Real-time traffic flow prediction made possible by massive traffic data and machine learning has helped people avoid gridlock by selecting less congested routes. To forecast or lower the danger of traffic accidents, big traffic data and machine learning may also offer a promising option (WHO, 2022).

# Research question, aims & objectives

We look at open datasets and other data sources related to this field to try to accurately predict risk levels (RTAs).

**Research question**

Can I predict road traffic accident risk areas in UK?

## Objectives

The following goals pertain to this work:

* Conduct a literature review to understand the existing work.
* To identify the factors that could have an impact on machine learning, use exploratory data analysis.
* Classify dangerous areas using spatial statistics on the secondary dataset.
* Using support vector machine methods, logistic regression, k-nearest neighbour, Rainforest, gradient boosting, and neural networks on a secondary dataset containing RTA characteristics These ML classifiers are employed mostly due to their distinguishing features and popularity in the literature.
* Evaluate and compare the performance of the classification models using various widely used evaluation techniques, such as accuracy, precision, recall, and the F1 score.
* To predict new regions, the best ML classifier is used.

## Deliverable

To investigate how localization patterns and hot spot distribution are used with temporal data, and to construct a highly precise machine learning model for predicting the likelihood of traffic accidents based on the spatiotemporal correlation pattern.

Literature review

This section analyses the literature within the domain that will assist to identify and justify the proposed project. The literature review focuses on the present research on road accident prediction and different models used to achieve it, it addresses the gaps to implement new solutions to the research problem.

One can get a general idea of the danger areas based on historical statistics of traffic accidents that have been gathered throughout the years. The risk distribution, however, varied significantly depending on the hour of the day, day of the week, and month of the year. Complex elements, such as crowd density, traffic flow, weather, events, etc., can also have an impact on accident risk. A fine-grained and dynamic prediction of accident risk cannot be made using historical statistics. Therefore, it is crucial to incorporate machine learning technology into traffic accident risk prediction to forecast the dynamic accident risk change in a precise and quantitative manner. Chen uses the Stack Denoising Autoencoder's human mobility properties to extrapolate the risk of traffic accidents in Japan (Liu et al., 2018) They did not, however, consider the cyclical and spatial distribution patterns of traffic accidents. The deficiencies in road engineering, crash investigation and post-crash management, legislative and driver education makes roads to be unsafe. Globally, road accidents are the main cause of death per year, reducing millions of lives on a yearly basis. Hence, a system that can predict the occurrence of traffic accidents or accident-prone areas can potentially save lives. The people don’t have to depend on the government to implement traffic engineering etc, they take ownership of the likeliness of accidents to occur and will be able to mitigate it which will save lives and money a company always must incur due to loss of life and goods,

(Ogwueleka, Misra, Ogwueleka, & Fernandez-Sanz, 2014)used Artificial Neural Networks able to relate input with output, enable large number of variables and can tolerate error, ANN provide solution to non-algorithm problems, able to use historical data to deliver fresh solutions, while this model provides effective prediction of road accidents its inability to identify accidents hot spots for better management of drivers along the road was not considered by the authors.

(Oyetunji, Oladeji, Falana, & Idowu, 2017)used naïve bayes a machine learning classifier as an effective prediction model for road accidents in Nigeria, the author demonstrated how the model classified each accident, weather was not included in the factors for prediction, not all states of the country were used for experiment also there is need to use a mechanism that can identify accident hot spots.

The authors (Al Najada & Mahgoub, 2016)implemented combined methodologies of a big data analytics architecture, machine learning classifiers, this system will be more feasibility on locations that utilize road engineering because of the use of road segments in the experiments, hence this experiment mayn’t work well in developing countries, the authors used the different machine learning for different purposes to which identify patterns and accidents hot spots was not included which is more feasible in a developing country.

The precision and recall metric were used to confirm accuracy of the prediction model due to data imbalance as against most authors who used the accuracy metric in machine learning classifier algorithm for prediction, (Hébert, Guédon, Glatard, & Jaumard, 2019) the author recognises the effect of data imbalance, but weather was not used as a factor for prediction, also the author didn’t recognise patterns that could identify accident hot spots.

# Research Design

**Research Philosophy**

The term philosophy of research refers to a system of belief and assumptions on knowledge development, this research project tends to provide a practical approach which informs future practise to a problem or gap identified through review of various literature, this type of research philosophy is called Pragmatism, hence this research philosophy will be applied in the methodologies and techniques (Saunders, Lewis, & Thornhill, 2007).

**Research Approach**

We have identified our research problem which is the research question and have adopted practical solutions or algorithms for improving the efficiency of the research problem and collect data to explore and solve it. This type of research approach is a mixed method of qualitative and quantitative research (Park et al., 2016). Hence this project is centred towards a mixed method of qualitative and quantitative research**.**

**Methodology**

The three main areas of this project (EDA, Spatio-temporal Analysis and Machine Learning) will require a combined and independent evaluation of the research goals. Another methodology to collect and analyse data is required for each of those evaluations: - EDA: Both qualitative and quantitative data can be used, also qualitative data can be transformed to quantitative data for data analysis which makes it a mixed-method methodology. Machine Learning The nature of machine learning means that models are almost fully evaluated on quantitative data, in particular the precision and processing time percentages. A simple comparison of results from various ML models will allow for a fair evaluation while controlling all possible external variables, such as the use of the same data set.

Machine Learning Model: machine learning algorithm is considered for predictive modelling. In other words, by applying a function f to the independent variable x we can predict the value of a dependent variable y.

Generally, to forecast future outcomes, we use predictive modelling or analytics. We collect past data known as historical data on an event for this purpose systematically. We train a statistical model afterwards. To forecast future results, we use a trained statistical model. There are a variety of classifiers used for Machine Learning, the classifiers to be used are support vector machine methods, logistic regression, k-nearest neighbour, Rainforest, gradient boosting, and neural networks. Support vector machine produces considerable precision and less calculating capacity, and it can be used for both regression and classification tasks, as an abbreviation of SVM. However, it is widely used for classification purposes. Rainforest is mainly used for large dataset as an algorithm for constructing decision trees. Naïve Bayes is a machine learning algorithm used to resolve classification issues, it is mainly used in text classification, which includes a large data set for training. As the requirements of the Machine Learning model are developed, further research will be carried to determine if more suitable technologies are available.

**Time Horizon**

The time horizon will be cross-sectional because the project is conducted within 3months.

**Data Collection and Data Analysis Methodology**

**Data Source**

I obtained data for Great Britain from 1979 on reported road crashes broken down by severity, quantity, and rates of recorded road casualties by category of road users (Department for Transport, 2020). London was chosen because of its commitment to increasing traffic safety, as did Transport for London. An extensive consultation by Transport for London led to the London Assembly's evaluation of the Transport Strategy in 2018 and the revision in 2022. (TfL). The streets of London, where people live, work, and spend their free time, can be shaped by how they get around, including through using the train, bus, and tube systems (Copyright, Greater London Authority, 2018).

The names of London boroughs and LSOAs are both included in the second dataset. The LSOA Atlas provides an overview of demographic and related statistics for each Lower Super Output Area (LSOA) in Greater London. The following is a list of the boroughs that make up London: The City of London, Barking and Dagenham, Barnet, Bexley, Brent, Bromley, Camden, Croydon, Ealing, Enfield, Greenwich, Hackney, Hammersmith and Fulham, Haringey, Harrow, Havering, Hillingdon, Hounslow, Islington, Kensington and Chelsea, Kingston upon Thames, Lambeth, Lewisham, Merton, Newham, Redbridge, Richmond upon Thames, Southwark, Sutton, Tower Hamlets, Waltham Forest, Wandsworth, Westminster (LSOA)Greater London Authority (GLA), 2014

**Spatial Analysis**

We can summarise the distribution of a variety of phenomena using spatial statistics. As a result, it aids in our decision-making. The use of spatial statistical analyses, including methods for pattern recognition, distribution analysis, and geographical correlations, allows the direct integration of spatial variables with conventional non-spatial statistical methods (neighbourhood, connections, and spatial relationships). spatial autocorrelation using the Global Moran's I tool, as well as Getis-OrdGi-based analysis of hot regions for all incidents. Arc GIS 10.2 and its add-ons will be used for all spatial processing.

# Ethics, RISKS, and Issues

|  |  |  |
| --- | --- | --- |
| **Potential Risks** | Issues | Mitigations |
| **Not securing Access to a Specialist Data Analysis and ML Computer system** | Existing Computer will not be able to perform Data Analysis and ML due to Large Dataset, query etc | Access to any cloud infrastructure like AWS etc before the start of the project. |
| **Not securing access to the Arc GIS Pro 10.2** | Approval of securing access to Arc GIS Pro 10.2 throughout the duration of the project | Initiation to be made at an early stage, subscription to be secured. |
| **Saved data loss** | Redo of the project | Cloud backup of files |

Graphical user interface

Description automatically generated with medium confidence

Gantt Chart

Timeline

Description automatically generated with medium confidence

Chart, waterfall chart

Description automatically generated

Timeline

Description automatically generated

References

Abdulkabir, M., Tunde, R. S., & Edem, U. A. (2015). Trend analysis on road traffic accident in nigeria.*Science Innovation, 3*(5), 52-57.

Al Najada, H., & Mahgoub, I. (2016). Anticipation and alert system of congestion and accidents in VANET using big data analysis for intelligent transportation systems. Paper presented at the *2016 IEEE Symposium Series on Computational Intelligence (SSCI),*1-8.

Hébert, A., Guédon, T., Glatard, T., & Jaumard, B. (2019). High-resolution road vehicle collision prediction for the city of montreal. Paper presented at the *2019 IEEE International Conference on Big Data (Big Data),*1804-1813.

Ihueze, C. C., & Onwurah, U. O. (2018). Road traffic accidents prediction modelling: An analysis of anambra state, nigeria.*Accident Analysis & Prevention, 112*, 21-29.

Ogwueleka, F. N., Misra, S., Ogwueleka, T. C., & Fernandez-Sanz, L. (2014). An artificial neural network model for road accident prediction: A case study of a developing country.*Acta Polytechnica Hungarica, 11*(5), 177-197.

Oyetunji, M. O., Oladeji, F. A., Falana, O. J., & Idowu, P. A. (2017). Prediction of road traffic accident in nigeria using naive baye’s approach.*Advances in Multidisciplinary & Scientific Research Journal, 3*(1), 23-30.

Ren, H., Song, Y., Wang, J., Hu, Y., & Lei, J. (2018). A deep learning approach to the citywide traffic accident risk prediction. Paper presented at the *2018 21st International Conference on Intelligent Transportation Systems (ITSC),*3346-3351.

Ukoji, V. N. (2014). Trends and patterns of fatal road accidents in nigeria (2006–2014).*Internet: Http://Www.Ifra-Nigeria.Org/IMG/Pdf/Fatalroad-Accidents-Nigeria.Pdf Nov, 28*

World Health Organization. (2018). *Global status report on road safety  
2018.* (). Retrieved from <https://www.who.int/publications/i/item/9789241565684>

World Health Organization. (2021). Road traffic injuries. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>

ppendix A: Completed Research Ethics FORM

Text

Description automatically generated

**UREC 1 RESEARCH ETHICS REVIEW FOR STUDENT RESEARCH WITH NO HUMAN PARTICIPANTS OR DIRECT COLLECTION OF HUMAN TISSUES, OR BODILY FLUIDS.**

All University research is required to undergo ethical scrutiny to comply with UK law. The SHU [Research Ethics Policy](https://www.shu.ac.uk/research/quality/ethics-and-integrity/ethics-policies) should be consulted before completing the form. Answering the questions below will confirm that the study fits this category and that any necessary approvals or safety risk assessments are in place. The supervisor will approve the study, but it may also be reviewed by the College Teaching Programme Research Ethics Committee (CTPREC) as part of the quality assurance process.

The final responsibility for ensuring that ethical research practices are followed rests with the supervisor for student research.

Note that students and staff are responsible for making suitable arrangements to ensure compliance with the General Data Protection Regulations (GDPR), for keeping data secure and if relevant, for keeping the identity of participants anonymous. They are also responsible for following SHU guidelines about data encryption and research data management. Information on the [ethics website](https://www.shu.ac.uk/research/quality/ethics-and-integrity/guidance-and-legislation)

The form also enables the University and College to keep a record confirming that research conducted has been subjected to ethical scrutiny.

The form may be completed by the student and the supervisor and/or module leader (as applicable). In all cases, it should be counter-signed by the supervisor and/or module leader, and kept as a record showing that ethical scrutiny has occurred. Students should retain a copy for inclusion in the appendices of their research projects, and a copy should be uploaded to the module Blackboard site for checking.

Please note if it may be necessary to conduct a health and safety risk assessment for the proposed research. Further information can be obtained from the Safety Co-ordinator.

**1. General Details**

|  |  |
| --- | --- |
| Name of student | Rita Uzoka |
| SHU email address | C0034087@my.shu.ac.uk |
| Course or qualification (student) | Msc. Big Data Analytics |
| Name of supervisor |  |
| email address |  |
| Title of proposed research | Road Traffic Accident Risk Prediction in the UK |
| Proposed start date | October 3,2022 |
| Proposed end date | January 9,2023 |
| Brief outline of research to include, rationale & aims (250-500 words). | Building a reliable method for predicting traffic accidents is a crucial step in the avoidance of traffic accidents. We can advise surrounding cars to warn them or prompt them to take a less risky route if the likelihood of a traffic collision in a specific area can be forecast. However, due to the multiplicity of elements that might influence traffic accidents, it is highly challenging to predict with accuracy the risk of an accident. For instance, the number of road accidents varies greatly between regions. Furthermore, unfavorable weather conditions like snow or fog can lower road visibility and traffic capacity, which raises the risk of traffic accidents. The number of traffic accidents varies throughout the day depending on the time of day, presumably reflecting the health of the drivers. Although a lot of study has been done on the primary components that contribute to traffic accidents. it is still difficult to estimate the likelihood of accidents occurring dynamically. |

I confirm that this study does not involve collecting data from human participants \_\_ (please tick)

**2. Research in Organisations**

|  |  |
| --- | --- |
| **Question** | **Yes/No** |
| 1. Will the research involve working with/within an organisation (e.g. school, business, charity, museum, government department, international agency, etc.)? | No |
| 1. If you answered YES to question 1, do you have granted access to conduct the research?   *If YES, students please show evidence to your supervisor. PI should retain safely.* |  |
| 1. If you answered NO to question 2, is it because:    1. you have not yet asked    2. you have asked and not yet received an answer    3. you have asked and been refused access.   *Note: You will only be able to start the research when you have been granted access.* |  |

**3. Research with Products and Artefacts**

|  |  |
| --- | --- |
| **Question** | **Yes/No** |
| 1. Will the research involve working with copyrighted documents, films, broadcasts, photographs, artworks, designs, products, programmes, databases, networks, processes, existing datasets or secure data? | No |
| 2. If you answered YES to question 1, are the materials you intend to use in the public domain?  *Notes: ‘In the public domain’ does not mean the same thing as ‘publicly accessible’.*   * *Information which is 'in the public domain' is no longer protected by copyright (i.e. copyright has either expired or been waived) and can be used without permission.* * *Information which is 'publicly accessible' (e.g. TV broadcasts, websites, artworks, newspapers) is available for anyone to consult/view. It is still protected by copyright even if there is no copyright notice. In UK law, copyright protection is automatic and does not require a copyright statement, although it is always good practice to provide one. It is necessary to check the terms and conditions of use to find out exactly how the material may be reused etc.*   *If you answered YES to question 1, be aware that you may need to consider other ethics codes. For example, when conducting Internet research, consult the code of the Association of Internet Researchers; for educational research, consult the Code of Ethics of the British Educational Research Association.* | The information is publicly accessible. |
| 3. If you answered NO to question 2, do you have explicit permission to use these materials as data?  *If YES, please show evidence to your supervisor.* | Yes |
| 4. If you answered NO to question 3, is it because:  A. you have not yet asked permission  B. you have asked and not yet received and answer  C. you have asked and been refused access.  *Note You will only be able to start the research when you have been granted permission to use the specified material.* | **NA** |

**Adherence to SHU policy and procedures**

|  |  |
| --- | --- |
| **Personal statement** | |
| I can confirm that:   * I have read the Sheffield Hallam University Research Ethics Policy and Procedures * I agree to abide by its principles. | |
| **Student** | |
| Name: Rita Chinwenwa Uzoka | Date:13/09/2021 |
| Signature: Rita | |
| **Supervisor or other person giving ethical sign-off** | |
| I can confirm that completion of this form has confirmed that this research does not involve human participants. The research will not commence until any approvals required under Sections 3 & 4 have been received and any health and safety measures are in place. | |
| Name: | Date: |
| Signature: | |
| Additional Signature if required: | |
| Name: | Date: |
| Signature: | |

**Please ensure the following are included with this form if applicable, tick box to indicate:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Yes** | **No** | **N/A** |
| Research proposal if prepared previously |  |  |  |
| Any associated materials (e.g. posters, letters, etc.) |  |  |  |
| Health and Safety Project Safety Plan for Procedures |  |  |  |