

Transcript: Research Proposal Presentation

Hallo, my name is Simbiso Makunde and I am studying the online Master of Science in Data Science program at the University of Essex. This module is called Research Methods and Professional Practice and I shall present a Research Proposal for a project titled "Implementing Machine Learning tools and/or techniques in Crime Prediction."

The points that will be covered in this presentation include a short description of the use of Machine Learning in Crime Prediction, the aims and objectives of the project and an analysis of the key literature related to the research question. Next, the Research design or Methodology shall be outlined, and then the Ethical consideration and risk assessment will be discussed. Last but not least, a description of the artefacts to be created and a timeline of the proposed activities will be presented.

Crime has been on the rise in recent years, and according to a 2024 report by the European Commission, police recorded robberies in the European Union increased by 9.7% between 2021 and 2022. In 2024 BBC reported that crime experienced by individuals and households in England and Wales increased by 10% over the previous year. As a result, the research into and development of crime prediction algorithms such as PredPol, CrimeRader by the United Nations Office on Drugs and Crime and BlueCRUSH are pivotal for various stakeholders in law enforcement strategies.

A study by Mohler in 2015 showed that a machine learning forecasting model can predict 1.4 to 2.2 times as much crime as a crime analyst, which means that by using Machine Learning in crime prediction, police patrols can apply their resources more efficiently in crime hotspots. As reflected in Alsaedi's paper of 2017, Machine Learning models can be applied in monitoring surveillance videos and social media platforms such as Twitter, to detect anomalous or criminal behaviour and threats of dangerous activity before they occur.

The research question for this project is "How can Machine learning tools and algorithms be ethically and effectively implemented to improve the accuracy and efficiency of predictive policing?"

The aims and objectives are: To review the existing literature on the topic, as well as to explore the machine learning tools and techniques which are currently being used in crime prediction. To collect, pre-process and analyse historical crime datasets, and lastly to develop and evaluate the performance of various prediction models.

Some of the related key literature include the 2023 study by Amin et al titled "Detection of Anomaly in Surveillance Videos using Quantum Convolutional Neural Networks" in which the authors demonstrate the potential of combining Convolutional Neural Networks and quantum computing to improve the accuracy of a classification algorithm in classifying anomalous activity in real-time surveillance systems. They

proposed two models for this classification task, that is a Javaeria Deep Convolutional Neural Network or DCNN , which had an accuracy of 99% and a Javaeria Quantum quantum Convolutional Neural Network or QCNN which had an accuracy of 97% .

Another key literature is the 2021 paper by Azwad et al titled “Crime Prediction and Forecasting using Machine Learning Algorithms” in which various models such as Random Forests, Neural Network, K-Nearest-Neighbours and AdaBoost were developed with the historical crime records of the Chicago Police Department in order to identify crime hotspots in in large cities. The study determined that the Neural Network-based model performed the best with an accuracy of 90.77%.

A study titled “Crime Prediction Model using Deep Neural Networks” was aimed at forecasting whether or not an individual is likely to commit crime in future years given historical records and also what what level of crime is likely to be committed. The authors used a Deep Neural Network using fully connected convolution layers to build the prediction model and it achieved 94% accuracy in predicting the level of crime to be committed (Chun et al, 2019).

The 2019 study by Mohler titled “Randomized Controlled Field Trials of Predictive Policing” is important as it highlights one of the biggest risks in using Machine Learning for predictive policing, which is that there can be biases in the training datasets and this can result in stigmatisation of certain individuals in society.

While these are some of the key literatures which are relevant to this research, there are other sources such as government reports and websites of official institutions which describe the current research and state of predictive policing.

The development strategy for an effective crime prediction Machine Learning model begins with Data Collection of historical crime datasets from sources such as governmental and law enforcement department websites. These datasets have quantitative crime data with attributes such as crime location, time of the crime and the nature of the crime. Once the appropriate dataset for a specific city has been selected, the data shall be pre-processed for statistical analysis i.e. Data Cleaning whereby missing values, outliers and duplicates are properly handled, Normalisation and Feature Extraction.

The next step is the development and implementation of various Machine Learning Models such as SVM, Random Forests, Neural Networks and Clustering algorithms. These models can be developed with Python-based libraries such as scikit-learn, and then the performance of the models can be evaluated with metrics such as accuracy and the F1 Score. With this development strategy, the most effective classification and crime hotspot forecasting algorithm can be determined.

Implementing Machine Learning tools and techniques in crime prediction raises some ethical concerns and there are also risks inherent in the process. Barocas and Selbst state in their 2016 paper that ML models that are trained on crime data

perpetuate existing Bias and discrimination in law enforcement measures, and as a result specific specific demographics are usually disproportionately targeted .This can be combated by selecting training and testing datasets which are representative of diverse populations.

Another important issue is Data Privacy and Security, and data professionals must ensure that the research design complies with data protection legal frameworks such as GDPR.

Accountability in decision making must be clear to the public and according to Olaye & Egon in 2024, providing clear information about how predictions are made, and how the law enforcement agencies use the data in predictive policing can help to ensure transparency and build public trust.

The artefacts which will be created in this research project include: a predictive model which forecasts crime locations, and the type of crime likely to be committed as well as high risk days or times. I shall also create a dashboard for crime monitoring with various visualisations such as a geographical representation of crime hotspots and time- series visualisations of the occurrences of crime. Last but not least, a technical report summarizing the dataset details, performance or evaluation metrics, as well as the ethical guidelines and the limitations of the crime prediction model will be among the artefacts to be created.

Here is a timeline for the proposed activities. The project is planned over the course of at least 4 months, and the timeline includes collecting data, developing the models, testing and validating them and lastly the reporting and creation of the documentation.

The data will be sourced from publicly available datasets, and pre-processing will last for approximately 3 weeks. Pre-processing shall involve normalizing features, encoding variables and handling missing values in the data in preparation for the Machine learning Models. This will be followed by Exploratory Data Analysis , which will last for about 2 weeks.

The next step is the selection and development of the suitable machine learning models, which will last for about 4 weeks. In this stage techniques such as feature engineering and hyperparameter tuning will be employed to enhance the predictive ability of the selected models.

This will be followed by the evaluation of the machine learning models with various performance metrics and Validation Methods. This process is expected to last for about 3 weeks.

The last stage is the Documentation and compilation of the Research Report which will include the Methodology, the findings as well as a description of the functionality and limitations of the proposed model. This should take about 3 weeks to finalize.

So now this is the end of my presentation. Thank you very much for listening to this presentation on my proposed Research Design. Please find the list of references at the end of the presentation. Thank you very much.

References:

Ae Chun, S. et al. (2019). Crime Prediction Model using Deep Neural Networks. *Proceedings of the 20th Annual International Conference on Digital Government Research*, 512-514, DOI:[10.1145/3325112.3328221](https://doi.org/10.1145/3325112.3328221)

Alsaedi, N. et al. (2017). Can We Predict a Riot? Disruptive Event Detection Using Twitter. *ACM Transactions on Internet Technology*. 17(2). 1-26.
DOI:[10.1145/2996183](https://doi.org/10.1145/2996183)

Amin, J. et al. (2023). Detection of anomaly in surveillance videos using quantum convolutional neural networks. *Image and Vision Computing*. 135, 1-12.
DOI:[10.1016/j.imavis.2023.104710](https://doi.org/10.1016/j.imavis.2023.104710)

Azwad, T. et al. (2021). Crime Prediction and Forecasting using Machine Learning Algorithm. *International Journal of Computer Science and Information Technologies*, 12(2), 26-33.

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, 104, 671-732. DOI: <https://dx.doi.org/10.2139/ssrn.2477899>

BBC (2024). *Crime up 10% over past year in England and Wales, ONS says.*

Available at <https://www.bbc.com/news/articles/cpwr89rv9qno> [Accessed: 27

December 2024]

Eurostat (2024). *Crime Statistics*. Available at https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Crime_statistics#:~:text=Police%2Drecorded%20crimes%20against%20property,compared%20with%20the%20previous%20year. [Accessed:

22 December 2024]

Mohler, G. O. et al. (2015). Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association*, 110(512), 1399–1411.

DOI:[10.1080/01621459.2015.1077710](https://doi.org/10.1080/01621459.2015.1077710)

Olaoye, F. & Egon, A. (2024). Predictive Policing and Crime Prevention. *Crime & Delinquency*. Available at:

https://www.researchgate.net/publication/383565286_Predictive_Policing_and_Crime_Prevention [Accessed: 29December 2024]