Research Proposal – Crime Analysis

Christopher Final

MSc Data Science

Agenda

01. Research Title and Question

02. Aims/objectives

03. Key Literature and Ethical Consideration

04. Methodology

05. Timeline for Project

Leveraging Data Science Techniques for Crime Data Analysis: Insights into Trends, Patterns, and Predictive Modelling

For my research project, crime data provides the opportunity to apply data science techniques such as:

- Trend analysis
- Prediction modelling
- Natural Language Programming (NLP)
- Clustering
- Recurrent neural networks (RNN)
- Geospatial analysis

Contribution to Crime Data Analysis

Recent advancements in crime data collection have sparked interest among data scientists and facilitated the integration of artificial intelligence (AI) in crime analysis. This progress allows for a more mixed-methods approach, moving beyond traditional qualitative methods of crime observation and review. An evidence-based approach can effectively guide policymakers in implementing efficient crime prevention strategies.

These strategies may help identify:

- **High-crime areas** - pinpointing locations with elevated crime rates to allocate resources effectively.
- **Determinants of crime** - analysing factors such as socio-economic conditions that contribute to crime rates.
- **Crime trends** - recognizing patterns over time, including seasonal variations in crime rates.

Research Problem

The Home Office (2018) estimated that the total cost of crime in England and Wales for the 2015/16 fiscal year amounted to around £50 billion for crimes committed against individuals, while crimes targeting businesses accounted for approximately £9 billion. This data highlights the significant economic impact of crime on both personal and commercial sectors within the region.

Whilst UK police forces have access to a vast amount of data, there is a lack in the technological capability to use it effectively (West Midlands Police and Crime Commissioner 2021).

Tools for analysing crime data and predicting future events have historically lagged behind, often struggling to keep pace with the rapid evolution of technology and the increasing complexity of criminal behaviour.

Crime has become more complex in modern times, influenced by several key factors:

- Digital crime
- Social media influences
- Economic stressors

The demand for advanced, data-driven approaches has increased, including:

- Use of artificial intelligence (AI)
- Implementation of machine learning technologies

Challenges faced by policing bodies in adopting modern tools:

- Limited budgets
- Inadequate technological infrastructure
- Insufficient training in data science methods

Research Questions

Primary Research Question

“How can advanced data science techniques, including machine learning and statistical modelling, be applied to crime data to detect patterns, understand contributing factors, and predict future crime occurrences?”

Secondary Research Questions

“How do socioeconomic variables, such as unemployment or income levels, correlate with crime patterns?”

“Can machine learning algorithms accurately predict crime hotspots?”

“What spatial (physical spaces) and temporal (cyclical) factors contribute to the rise in crime in certain areas?”

Aims and objectives

Aims -

- To explore and implement various data science techniques for the analysis of crime data.
- To build predictive models that can anticipate future criminal activity based on historical data.
- Visualise data to create compelling story telling to aid decision-making.

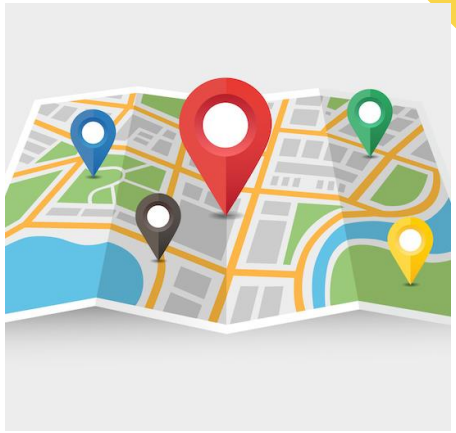
Achieving the aims of the research project will involve –

- **Data Collection** - Collecting crime datasets for manipulation and interpretation (e.g. data.police.uk).
- **Key Factors** - Identify key factors and features that influence crime patterns (e.g., geographic location, demographic factors, social environment).
- **Machine Learning Techniques** - Apply machine learning models such as classification, clustering, and time series analysis to detect patterns.
- **Visualise and Prediction Accuracy** - Display findings using visualisations while providing analysis on the predictive accuracy and performance of the models implemented.

Key Literature

Gerber, M. S. (2014)

Predicting Crime Using Twitter and Machine Learning.



Valasik, M.(2024)

Crime Mapping and Spatial Analysis.

Khan et al (2019)

Predictive policing: A machine learning approach to predict and control crimes in metropolitan cities.

Gerber, M. S. (2014). "Predicting Crime Using Twitter and Machine Learning."

Gerber (2014) explores the potential of social media as a tool for predicting crime, highlighting its influence when combined with Machine Learning techniques. He discusses how models like decision trees, support vector machines, Natural language programming (NLP), logistic regression, and random forests can be applied to crime data to uncover patterns and trends.

However, Gerber also emphasises the critical need for data accuracy and the avoidance of bias to ensure that these models do not yield discriminatory results.

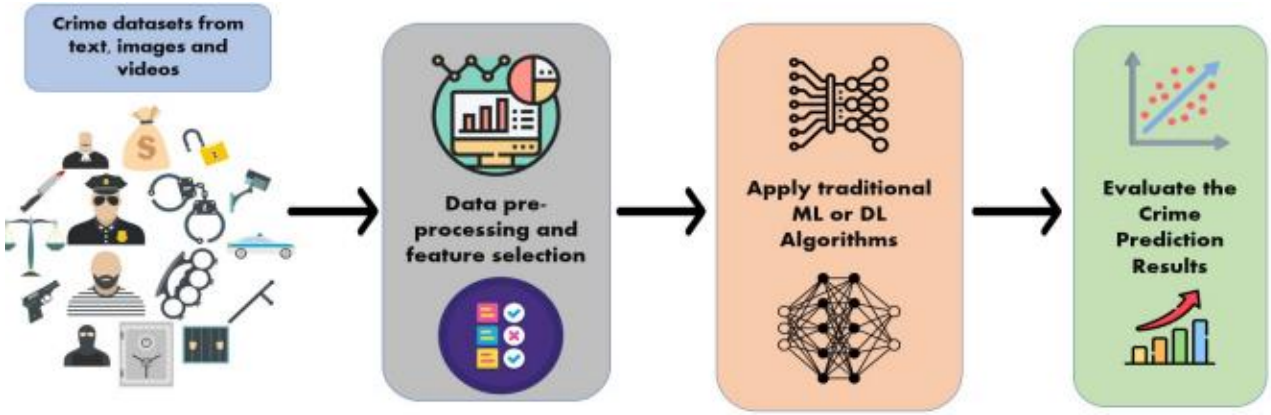


FIGURE 11. Architecture flow of crime prediction

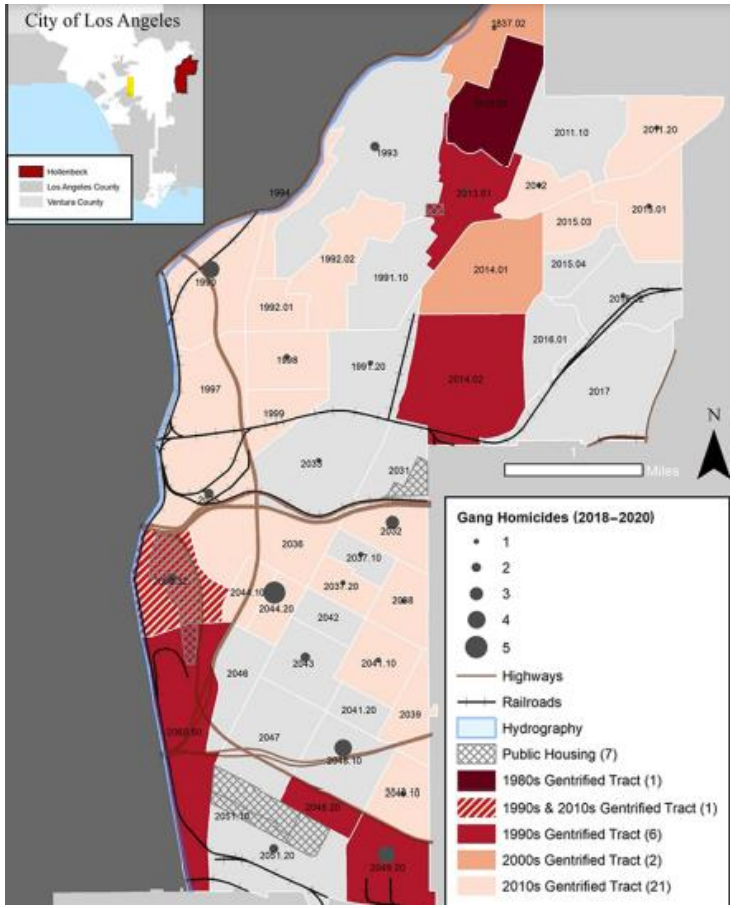
FIGURE 7. Word cloud on selected articles

Valasik, M. (2024) Crime Mapping and Spatial Analysis.

Geographic information gives the opportunity to visualise spatial patterns, allowing for analysis like multivariate regression (Valasik 2024).

The use of exploratory spatial data analysis (ESDA) in the study of American neighbourhoods was used to visualise spatial distributions, whilst also detecting clusters. ESDA allowed for the correlation between variable before statical modelling.

Valasik (2024) also mentions how spatial weights matrix can be used and the important of assigning weights when applying the matrix and visualising spatial dependences.



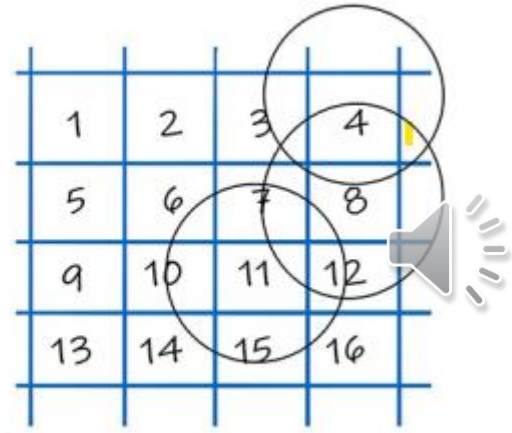


Figure 3. Spatial distribution of county poverty rates in 2012 using LISA statistics.

Khan et al (2019) Predictive policing: A machine learning approach to predict and control crimes in metropolitan cities.

Khan et al. (2019) discussed the application of machine learning techniques for predictive analysis.

One such technique, K-means clustering, is an intuitive and widely used method that iterates through data until the centroids stabilize, indicating a successful partition. It is particularly effective with large datasets.

Additionally, the Naïve Bayes classifier, another technique highlighted, calculates the probability of different categories (e.g., crime types) based on underlying data. This method is especially beneficial in text classification tasks.

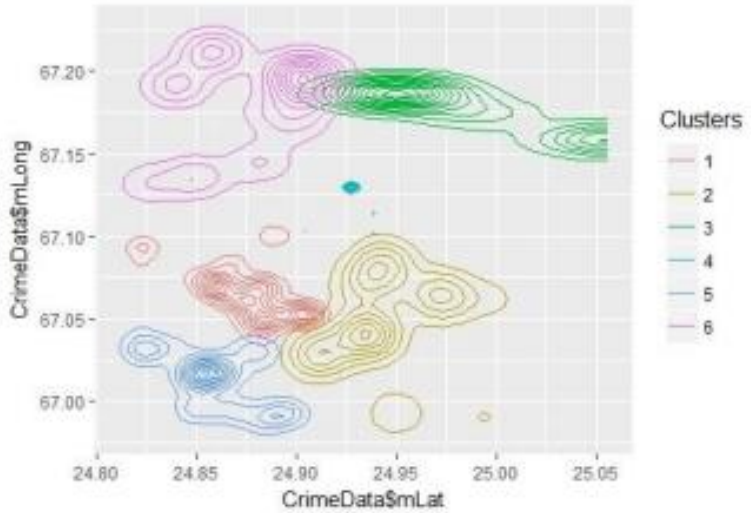


Figure 4. Crime density clusters

Ethics – Bias and Fairness

When implementing machine learning models, it is crucial to incorporate ethical considerations into the training process. As highlighted by Barocas and Narayanan (2019), ensuring fairness in machine learning is essential for preventing bias, and these models should not replace human decision-making.

This approach is particularly important to mitigate any disproportionate impacts on vulnerable communities. Such considerations align with broader discussions on the ethical use of algorithms in various domains, emphasizing the need for transparency and accountability in AI systems.

Methodology

For the effective modelling of crime analysis and prediction the following steps will be taken.

1. Data Collection
2. Data preprocessing
3. Modelling
4. Evaluation
5. Visualisation

Data Collection – Utilize open-source data sources, such as data.police.uk, to select a specific area for crime data analysis.

Data Preprocessing – Ensuring a clean dataset is crucial. Therefore, data from data.police.uk should be examined and standardized to apply scaling and normalisation techniques effectively.

Modelling – Techniques like K-means clustering can be employed to identify crime hotspots. Classification algorithms, such as Random Forest, will be utilised to predict the probability of future crimes. Additionally, time series analysis can be performed using recurrent neural networks, like Long Short-Term Memory (LSTM), to recognize complex patterns in sequential data.

Evaluation – Implement cross-validation techniques, where the dataset is divided into subsets, to evaluate the model's performance on unseen data, thereby helping to prevent overfitting.

Visualisation – Leverage libraries such as Plotly to create visualisations of crime trends, hotspots, and predictive models, enhancing storytelling for non-technical audiences.

Timeline for Research Project

Data Collection & Preprocessing (1 month)

Collecting and cleaning data for analysis and normilisation

Exploratory Data Analysis (1.5 months)

Gather basic understanding of the data and patterns. (e.g. clustering)

Modelling and Development (2 months)

Implement machine learning methodology to create predictions

Evaluation and Validation (1.5 months)

Evaluate models implemented on crime data and adjust where necessary.

Visualisation development (1 month)

Build and create visuals to showcase findings to potential stakeholders.

Final Analysis and Reporting (1 month)

Finalise analysis and key findings with visualisations for explaining key concepts.

Conclusion

The aim of the project is to help leverage the use of data science techniques, when analysing crime. Providing insights into trends, patterns, and future predictions.

The use of mixed methodology will help with achieving this. However, the importance of ethical considerations and the need for fairness in machine learning cannot be overstated, as they will guide my approach to ensuring that the insights gained do not perpetuate biases.

References

UK Government (2019) The economic and social costs of crime. Available at: https://www.gov.uk/government/publications/the-economic-and-social-costs-of-crime [Accessed: 1st October 2024].

West Midlands Police and Crime Commissioner (2021) National Data Analytics Solution (NDAS) Violence Case Study. Available at: https://www.westmidlands-pcc.gov.uk/wp-content/uploads/2022/01/2021-11-03-EC-Agenda-Item-2.1-NDAS-VC-National-Data-Analytics-Solution.pdf [Accessed: 1st October 2024].

Gerber, M.S. (2014) Predicting crime using Twitter and kernel density estimation. Decision Support Systems, 61, pp.115-125.

Valasik, M., 2024. Crime Mapping and Spatial Analysis. Oxford Research Encyclopaedia of Criminology. Available at: https://oxfordre.com/criminology/view/10.1093/acrefore/9780190264079.001.0001/acrefore-9780190264079-e-821 [Accessed: 5th October 2024].

Khan, J.R., Saeed, M., Siddiqui, F.A., Mahmood, N. and Arifeen, Q.U. (2019) Predictive policing: A machine learning approach to predict and control crimes in metropolitan cities. University of Sindh Journal of Information and Communication Technology, 3(1), pp.17-26.

Barocas, S., Hardt, M. and Narayanan, A. (2019) Fairness and Machine Learning. Available at: https://fairmlbook.org [Accessed: 6th October 2024].

Data.police.uk (n.d.) Police data for England and Wales. Available at: https://data.police.uk [Accessed: 6th October 2024].