PREDICTIVE POLICING: CRIME TRENDS IN LOS ANGELES

GROUP 17 – DATA MINING AND DATA ANALYTICS – CA683

S.NO	NAME	STUDENT ID	MAIL-ID
1	ALEN GEORGE SHIBU	23271013	alen.georgeshibu7@mail.dcu.ie
2	ALEENA JOSEPH	23264236	aleena.joseph6@mail.dcu.ie
3	POOJA ASHOK	23270600	Pooja.ashok2@mail.dcu.ie
4	JOSEPH MATHEW	23267632	joseph.mathew8@mail.dcu.ie
5	ABHIJITH HARIKUMAR	23269494	abhijith.harikumar2@mail.dcu.ie

DATASET LINK - https://www.kaggle.com/datasets/asaniczka/crimes-in-los-angeles-2020-2023

GITHUB LINK - https://gitlab.computing.dcu.ie/josepha6/ca683_los_angeles

ABSTRACT

This study, "Predictive Policing: Crime Trends in Los Angeles (2020-2023)," employs the CRISP-DM methodology to analyze and forecast crime trends in Los Angeles from 2020 to 2023. Using a dataset of over 850,000 crime incidents, the research progresses through the CRISP-DM phases - business understanding, data understanding, data preparation, modeling, evaluation, and deployment. By applying Random Forest for spatial analysis and LSTM networks for temporal forecasting, the study accurately identifies potential crime hotspots and trends. The findings, demonstrating an 83.93% accuracy in hotspot prediction, highlight the effectiveness of the CRISP-DM framework and data science in optimizing law enforcement strategies and enhancing public safety.

I. INTRODUCTION

In recent years, Los Angeles has witnessed evolving crime patterns, reflecting broader national trends and the unique socio-economic dynamics of the city. This evolution raises significant concerns for public safety and law enforcement, prompting the need for innovative approaches to crime prediction and prevention. Predictive policing, leveraging advanced data analytics and machine learning, offers promising pathways to understand and mitigate crime effectively. By analysing historical crime data, predictive models can identify potential hotspots and temporal patterns, enabling proactive law enforcement strategies.

The motivation behind this study is twofold: firstly, to harness the power of data analytics in enhancing public safety by predicting crime hotspots and trends with high accuracy; secondly, to contribute to the growing body of knowledge on the effectiveness of machine learning techniques in crime prediction, providing law enforcement agencies with actionable insights for resource allocation and strategic planning.

The primary research question guiding this study is: How have crime patterns in Los Angeles evolved from 2020 to 2023, and can we predict future hotspots by analysing these trends? To address this question, the objectives are to:

1. Analyse historical crime data from Los Angeles to identify patterns and trends from 2020 to 2023.
2. Utilize Random Forest Classifier and LSTM neural network models to predict future crime hotspots and temporal patterns.
3. Evaluate the accuracy and effectiveness of these predictive models in aiding law enforcement strategies.

Overview of Following Sections
Literature Review: Summarizes existing research on predictive policing and crime prediction methodologies, underscoring the contribution of this study to the field.

Methodology: Details the data pre-processing steps, feature selection, and the specifics of the Random Forest Classifier and LSTM models used for prediction.
Results: Presents the outcomes of the predictive models, including accuracy, F1 scores, and error metrics, alongside a discussion on the implications for law enforcement.
Conclusion and Future Work: Concludes with the study's implications for predictive policing practices and outlines potential directions for future research in enhancing crime prediction models.

II. RELATED WORK

The study [1] Proposed a predictive system integrating classification and visualization techniques to identify crime hotspots. Used publicly available crime data from a major city. Achieved over 90% accuracy in hotspot prediction. Suggested improving data granularity to enhance prediction accuracy.
Paper [2] Evaluated various classification algorithms for crime prediction, comparing their performance across multiple datasets. Utilized crime datasets from Chicago, Los Angeles, Egypt, and the United States. Found Random Forest to outperform other algorithms, with accuracy rates up to 92%. Recommended exploring hybrid models combining the strengths of various algorithms.

[3] Conducted a comprehensive analysis using data mining techniques for crime analysis and prediction. Analysed data from the National Crime Records Bureau. Highlighted the effectiveness of clustering, classification, and association rules in crime prediction. Pointed out the need for real-time crime data processing systems.
Applied machine learning and deep learning models to predict and forecast crime in two major cities. Focused on Chicago and Los Angeles crime statistics [4]. Demonstrated that XGBoost and LSTM models performed best for predictive accuracy and time series analysis, respectively. Suggested integrating more diverse datasets and real-time data for dynamic prediction models.
The Literature Review [5] Emphasized the use of advanced algorithms and investigative analysis tools for crime analysis. Employed a variety of datasets, including forensic data. Achieved high accuracy in crime prediction, underlining the potential of ID3 algorithm and hidden link detection. Encouraged the development of more intuitive tools for law enforcement agencies.
[6] Investigated the utility of geospatial analysis for predicting crime patterns and identifying emerging hotspots. Utilized geocoded crime incidents from an urban metropolitan area. Successfully identified spatial patterns and temporal trends, suggesting targeted policing could be more effective. Recommended incorporating socio-economic data to refine predictive accuracy.
Examined the impact of social media analysis on crime prediction,

focusing on sentiment analysis and predictive modelling [7]. Analysed social media posts alongside crime reports from urban areas. Found correlations between negative sentiment in social media posts and crime rates, indicating potential for early warning systems. Suggested deeper integration of natural language processing (NLP) techniques to enhance predictive models.

[8] Applied a hybrid model combining time series analysis and machine learning for crime forecasting. Focused on historical crime data over a decade in a large city. Demonstrated that combining ARIMA models with Random Forest significantly improved forecast accuracy over using either method alone. Proposed the exploration of neural network models for better handling of complex patterns in crime data.

III. DATA MINING METHODOLOGY

Our approach to analysing crime trends and predicting future hotspots in Los Angeles is rooted in the CRISP-DM framework, comprising six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. Each phase played a crucial role in guiding our analysis and ensuring the relevance and accuracy of our findings.

1. Business Understanding

The primary objective of this study was to understand the evolution of crime patterns in Los Angeles from 2020 to 2023 and to predict future crime hotspots. This goal was informed by the broader mission to enhance public safety through data-driven insights, providing law enforcement with actionable intelligence for pre-emptive measures.

2. Data Understanding

We utilized a dataset from Kaggle titled "Crimes in Los Angeles (2020-2023)," which contains 852,950 records of reported crimes, including details on the date, time, location, victim information, and crime status. An initial assessment revealed the dataset's comprehensiveness in capturing various aspects of crime incidents, making it suitable for our analysis.

3. Data Preparation

Data cleaning involved removing outliers, such as records with unrealistic victim age values, and refining attributes for clarity, such as splitting the 'Date_occurred' into separate date and time columns. Missing values in critical columns were retained to preserve the dataset's integrity. We also performed feature selection to identify relevant predictors for our models, focusing on demographic insights, incident specifics, and location details.

4. Modelling

We applied two primary modelling techniques:

1. Random Forest Classifier: To identify crime hotspots, we trained a Random Forest model using selected

features. The model's ability to handle high-dimensional data and its robustness against overfitting made it ideal for our analysis.

2. Long Short-Term Memory (LSTM) Neural Network: For temporal pattern analysis, an LSTM model was used to capture the sequences in crime occurrences over time, leveraging features like the day of the week and monthly crime counts.

5. Evaluation

The models were evaluated based on their accuracy, F1 score (for the Random Forest Classifier), and error metrics such as MAE, MSE, and RMSE (for the LSTM model). These metrics provided insights into the models' predictive capabilities and their practical applicability in a law enforcement context.

6. Deployment

While the deployment of models in a real-world setting falls beyond this study's scope, we discuss potential applications and considerations for integrating our findings into law enforcement workflows, emphasizing the need for continuous model retraining and validation against new data.

IV. EVALUATION/ RESULTS

1. Evaluation Methodology

a. Random Forest Classifier: The model's performance was evaluated using accuracy and F1 score. Accuracy provided a straightforward measure of the model's overall predictive capability, while the F1 score offered insights into the balance between precision and recall, crucial for applications in crime prediction where both false positives and false negatives carry significant implications.

b. LSTM Neural Network: For the LSTM model, we used Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) as our primary evaluation metrics. These metrics were chosen to quantify the model's prediction accuracy regarding temporal crime trends, with a lower score indicating higher accuracy.

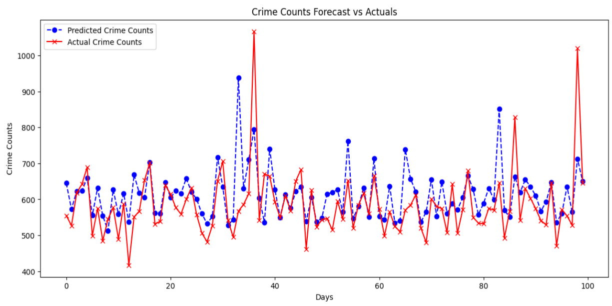

Fig.1. Crime Counts Forecast vs Actuals

2. Parameterization: For the Random Forest Classifier, we focused on the number of trees and their maximum depth, selecting values that prevented overfitting while allowing the model to adapt to diverse crime patterns. Similarly, the LSTM Neural Network's configuration, including the number of layers, units per layer, and dropout rate, was optimized to accurately capture temporal patterns in crime data without succumbing to overfitting, thanks to regularization techniques.

3. Results: The Random Forest Classifier achieved an accuracy of 83.93% and an F1 score of 0.8231, indicating a high level of

effectiveness in predicting crime hotspots. The LSTM model reported an MAE of 44.33, an MSE of 6117.80, and an RMSE of 78.22, demonstrating its capability to forecast crime trends with reasonable accuracy.

4. Discussion and Implications: Our study demonstrates that using data-driven models like the Random Forest and LSTM can effectively identify crime hotspots and predict crime trends in Los Angeles. These models provide law enforcement with valuable insights for strategic planning and resource allocation, enabling more targeted and proactive crime prevention strategies. This approach has the potential to improve public safety outcomes by making better use of resources.

V. CONCLUSION

Our study has effectively demonstrated the application of Random Forest and LSTM Neural Network models to identify crime hotspots and predict temporal crime trends in Los Angeles. These models have proven to be potent tools for enhancing law enforcement strategies, enabling proactive and targeted approaches to crime prevention. By analysing historical crime data with these advanced algorithms, we have provided insights that could lead to more efficient resource allocation and improved public safety outcomes.

VI. LIMITATIONS AND FUTURE WORK

The primary limitation of our study is its reliance on historical crime data, which may not fully reflect the dynamic nature of criminal behaviour and the impact of sudden societal shifts. This limitation points to a crucial area for future work: enhancing predictive accuracy by integrating a wider array of variables, including socio-economic factors and real-time data. Future research should also focus on developing hybrid models that merge various machine learning techniques for deeper insights into crime patterns. Additionally, applying these models in real-world law enforcement scenarios would offer invaluable feedback for further refining the approaches and better addressing the complexities of crime prevention strategies.

VII. REFERENCES

[1] S. Sathya Devan, D. M.S, and S. Gangadharan, "Crime Analysis and Prediction Using Data Mining," available at IEEE Xplore, https://ieeexplore.ieee.org/abstract/document/6906719.

[2] S. Samir and E. M. Mohamed, "Detecting and Predicting Crimes using Data Mining Techniques: Comparative Study," available at https://journals.ekb.eg/article_207749_6851eb18a1c11190f9444519b5fba5a0.pdf.

[3] S. Vijayarani, E. Suganya, C. Navya, and P. Student, "A Comprehensive Analysis of Crime Analysis Using Data Mining Techniques,":

https://www.ijcse.net/docs/IJCSE20-09-01-021.pdf.

[4] W. Safat, S. Asghar, and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," IEEE Xplore, : https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9424589.

[5] C. Chauhan and S. Sehgal, "A review: crime analysis using data mining techniques and algorithms," in IEEE Conference Publication, https://ieeexplore.ieee.org/document/8229823.

[6] T. Almanie, R. Mirza, and E. Lor, "Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 4, 2015, pp. 01-19, doi: https://doi.org/10.5121/ijdkp.2015.5401.

[7] S. M. Rajesh, I. Chiranmai, and N. Jayapandian, "Machine Learning Based Crime Identification System using Data Analytics," IEEE, 2023, doi: https://doi.org/10.1109/icscna58489.2023.10370717.

[8] "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data," IEEE Journals & Magazine, https://ieeexplore.ieee.org/document/8768367.

[9] "Crime Analysis Framework for Predicting Criminal Behavioral Patterns with Machine Learning," IEEE Conference Publication, https://ieeexplore.ieee.org/document/10236417.

[10] "Crime Data Analysis and Prediction Using Ensemble Learning," IEEE Conference Publication, : https://ieeexplore.ieee.org/document/8663186.

[11] "Crime Analysis Mapping, Intrusion Detection - Using Data Mining," IEEE Conference Publication, : https://ieeexplore.ieee.org/document/9140074.

[12] "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm," IEEE Conference Publication, : https://ieeexplore.ieee.org/document/8862691.