Grenze International Journal of Engineering and Technology, June Issue

Community Crime Analytics Dashboard using Machine Learning and Hotspot Detection

M Shilpa[1], H R Lithesh[2], Hari Om[3], Dhanush V[4] and Dhanush S[5]
[1]Associate Professor, Department Information Science Engineering, Bangalore Institute of Technology Bengaluru, India
Email: shilli22@yahoo.com
[2-5]Bachelor Engineering, Department Information Science Engineering, Bangalore Institute of Technology Bengaluru, India.
Email: hrlithesh28@gmail.com, dhanushvmodiliay@gmail.com, emeraldsilver007@gmail.com, dhanushgowda2108@gmail.com

***Abstract*— This paper outlines the design and features of a community-focused crime analytics dashboard. The system is designed to provide actionable safety insights to tourists and local citizens by leveraging machine learning and data visualization techniques. The platform analyzes historical crime data from sources like NCRB/Kaggle to visualize trends and distributions. A key feature is the integration of a simulated real-time crime feed via NewsAPI, which displays the latest news articles as live alerts with an associated confidence score. The predictive core of the dashboard uses an XGBoost model to generate a crime risk score or predict the most probable crime type for a user-specified location and time. All information is presented through an interactive Streamlit dashboard, which features mapbased hotspot visualizations using OpenStreetMap. The dashboard is designed with a simple, tabbed interface for ease of navigation, making complex data accessible to a general audience.**

***Index Terms*— Machine Learning, Hotspot Detection, Crime Analytics, XGBoost, Streamlit, Geospatial Visualization, Real-Time Data, OpenStreetMap.**

I. INTRODUCTION

Personal safety in urban environments is a growing concern for both residents and tourists. While vast amounts of crime data are collected, there is a significant gap in translating this data into accessible and easyto-understand insights for the general public. Existing tools are often designed for law enforcement and are not suitable for everyday use. This project proposes a Community Crime Analytics Dashboard to address this gap. The platform aims to empower users by providing a comprehensive safety tool that combines historical data analysis with simulated real-time updates. By visualizing crime trends and predicting potential risks, the dashboard enables users to assess the safety of an area at a specific time, helping them make more informed decisions.

A. Research Gaps

Although data-driven crime analysis is advancing, significant research gaps exist in creating effective public-facing safety platforms. Current predictive frameworks tend to rely heavily on static, historical datasets, which limits their adaptability to the dynamic, real-time nature of urban crime patterns. This creates a disconnect, as models trained on past data may fail to capture emerging, short-term threats.

Grenze ID: 01.GIJET.12.2.407_21

Furthermore, there is a gap in providing hyper-local and contextual information; city-level statistics are often too broad to be actionable for a tourist or local resident needing neighbourhood-specific advice. The challenge is compounded by the difficulty of integrating diverse data sources, balancing the accuracy of siloed official records with the timeliness of unverified real-time feeds like news or social media.

Another key gap lies in the usability and interpretability of these systems for a general audience. Most platforms are either complex analytical tools designed for law enforcement or overly simplistic apps that lack robust predictive power.

This leaves a need for a balanced system that is both analytically sound and intuitively designed for public use. Explainability and trust also remain unresolved challenges, as many systems provide "black-box" risk scores without justifying why an area is deemed dangerous, thereby reducing user confidence and actionable insight. Finally, there is a lack of personalization in safety guidance; current platforms rarely offer recommendations tailored to an individual's specific route, profile, or the time of day, making the advice generic rather than user-centric.

B. Trends and Applications

Recent trends in public safety highlight the integration of AI and machine learning to overcome the limitations of traditional, reactive methods. A major trend is the adoption of advanced geospatial analysis combined with real-time data streams from sources like news APIs, providing dynamic situational awareness. Platforms are increasingly embedding predictive models like XGBoost to forecast crime hotspots and using interactive dashboards built with frameworks like Streamlit to enhance transparency and user trust. These technologies are finding direct application in the tourism sector, where visitors can plan safer routes, and among urban residents, who can use these tools to make informed daily commute and housing decisions.

These advancements signify a broader shift towards public empowerment within smart city initiatives. By automating complex crime data analysis and offering personalized alerts, AI-powered dashboards democratize access to vital safety intelligence that was previously confined to law enforcement. This indicates a paradigm shift from a top-down, agency-focused model of public safety to a more accessible, scalable, and inclusive approach where every citizen is an informed participant in their own security.

C. Related work

Crime Type and Occurrence Prediction Using Naïve Bayes - One prominent approach utilizes Naïve Bayes variants on Kaggle data, emphasizing strong preprocessing like duplicate removal and feature mapping to model the independent effects of crime attributes. Its advantage lies in efficiently handling both categorical and numerical data to achieve high accuracy (93.07%) with transparent and interpretable outputs that support law enforcement decisions. However, a key drawback is its assumption of feature independence, which may overlook complex data relations, and its reliance on historical datasets that can have gaps or biases without incorporating real-time learning.

Crime Prediction Using KNN and SVM - Another study applied K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) to Indian crime data, incorporating geographic and social diversity factors after extensive data cleaning and feature selection. The innovation of this method is its ability to capture complex spatial patterns for effective hotspot detection and provide scalable solutions with alerts, achieving competitive accuracy of around 93%. The limitation, however, is its dependence on static historical datasets with no capacity for real-time adaptation, along with limited solutions for class imbalance and a lack of explicit focus on fairness and ethical considerations.

Urban Crime Prediction using Light Gradient Boosting Machine (LightGBM) - Research using San Francisco crime data employed LightGBM with Bayesian optimization for multi-class crime prediction, using feature engineering and visualizing results with heatmaps. This method's strength is that it achieved better accuracy than Naive Bayes, Random Forest, and even XGBoost, and successfully captured detailed temporal and spatial trends. The drawback is that its accuracy is limited by dataset size and sensitivity to noisy data, and it does not include broader socio-economic factors in its analysis.

Systematic Review of ML and DL for Crime Prediction - A comprehensive systematic review of over 150 articles categorized various ML and DL algorithms (e.g., decision trees, CNNs, RNNs) and identified key datasets and trends. Its value is in providing a broad overview of effective techniques and their applications in areas like predictive policing. However, the review highlights significant drawbacks in the field, including challenges with the quality and availability of crime data, the limited interpretability of complex models, and the fact that scalability and fairness issues remain unresolved in sparse real-world deployments.

D. Objectives

- ➢ To design and develop a community crime analytics dashboard integrating a machine learning model (XGBoost), simulated real-time data feeds (NewsAPI), and interactive geospatial visualization for effective public safety awareness.
- ➢ To implement a scalable web-based architecture using Python and Streamlit, ensuring an accessible, responsive, and user-friendly interface for citizens and tourists.
- ➢ To evaluate the platform's predicted effectiveness in enhancing public safety awareness, improving accessibility to crime data, and empowering users to make proactive decisions compared to the lack of existing public-centric tools.

E. Methodology

Figure.1 Illustrates the methodology flow of the Community Crime Analytics Dashboard. The process begins with comprehensive data collection and preprocessing of historical crime records, followed by the ingestion of real-time crime data. This prepared data then feeds into the machine learning prediction module. Finally, the processed insights and predictions are presented through an interactive dashboard to the end-user.

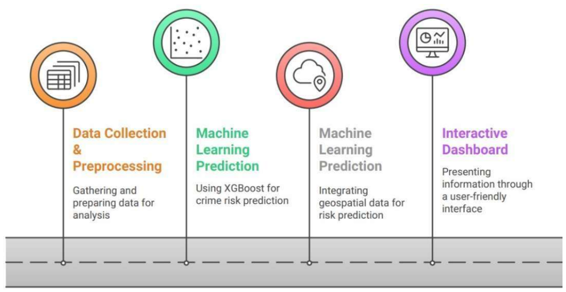

Figure 1 Methodology Flow of Community Crime Analytics Dashboard

F. Problem Statement

Many existing crime analysis tools only visualize historical data and cannot forecast future incidents. Law enforcement and citizens often face challenges in clearly identifying crime hotspots due to inadequate or complex visualization tools. Furthermore, there is no single, integrated system that effectively combines prediction, mapping, and data analytics to support the public's decision-making. A user-friendly, predictive, and integrated digital platform is needed to bridge this gap and empower citizens and tourists with proactive safety insights.

II. Existing System Flow

Figure.2. Illustrate the current process forces users to manually research fragmented online sources and interpret raw, outdated data. This leads to a purely subjective assessment of risk and an uncertain decision, as it lacks predictive analytics or integrated visualization.

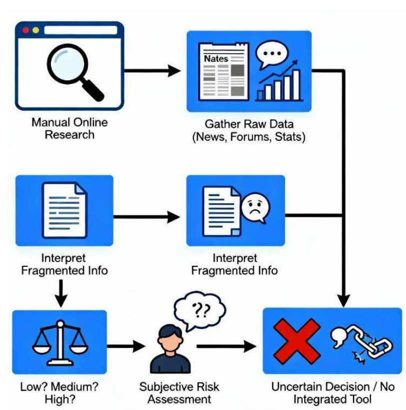

Figure 2 Existing System Flow Chart for Assessing Crime Safety

III. PROPOSED SYSTEM AND ARCHITECTURE

Figure.3 Represents the architecture of the dashboard, a web-based platform powered by Python and Machine Learning. It integrates an interactive map, a predictive analytics engine, a live crime feed, and historical trend charts to support tourists and local citizens. Features like color-coded hotspots, predictive risk scores based on time and location, and real-time alerts from news sources make it user-friendly. A dedicated section for visualizing historical data helps users understand long-term crime patterns in their city. The system thus empowers citizens with data-driven insights to enhance their personal safety and situational awareness.

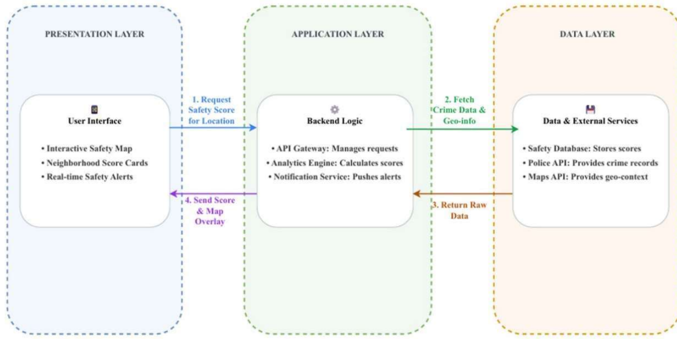

Figure 3 Proposed System Architecture

A. Use Case Representation

Figure.4 Illustrates the use case of the dashboard. The primary actor is the User (Tourist or Local Citizen), who interacts with the system to access safety scores, view crime hotspots, and receive real-time alerts. The platform integrates advanced tools such as an XGBoost model for prediction, a Streamlit frontend, NewsAPI for live data, and OpenStreetMap for geospatial visualization to deliver predictive safety analytics. Users gain access to interactive maps and risk assessments, allowing them to make informed decisions about their routes and destinations.

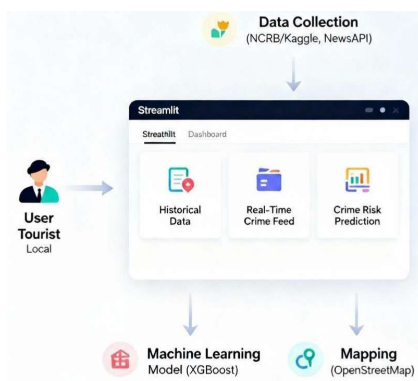

Figure 4 Use case Interaction and Data Flow

B. Risk Assessment Algorithm

A machine learning-assisted pipeline that analyzes historical and real-time data, maps crime hotspots, retrieves relevant temporal context, and presents a ranked risk score with a summary of recent incidents.

Step-by-step execution

Step-1: Collect and preprocess

Ingest historical crime data (NCRB/Kaggle) and live news feeds (NewsAPI); extract location and keywords, clean data, and perform feature engineering.

Step-2: Authenticate and sessionize

The user accesses the public Streamlit dashboard; the session begins for user-specific queries (e.g., location input).

Step-3: Analyze dialogue
Run the XGBoost model on user-provided inputs (location, time) and historical data to generate a predictive risk score and identify the most probable crime type.
Step-4: Retrieve knowledge
Query NewsAPI for recent, geo-tagged incidents in the specified area to provide up-to-the-minute context.
Step-5: Generate options
Produce a color-coded hotspot overlay on the OpenStreetMap and display the risk score, probable crime types, and recent news alerts on the dashboard.
Step-6: Negotiate loop
Present the map and alerts, capture new user inputs for different locations or times, and refine the visualization accordingly.
Step-7: Finalize and record
Render the final risk assessment, map visualization, and alert summary for the user's current query.

IV. Result and Discussion

Figure.5 This interactive dashboard view from showcases a dynamic crime heatmap overlaid on a Google Maps interface, centered around the user's current location. Red areas indicate high-risk crime hotspots, while green signifies safer zones. Clicking on a hotspot, such as "Market Chowk" or "Green Park Colony," reveals detailed predictive analytics, including the number of incidents over the last two years, specific high-risk timings, and the most common types of crimes likely to occur in that vicinity (e.g., "Pickpocketing, Chain Snatching, Minor Theft" or "Assault, Vandal"). On the left, a personalized "Your Current Risk Assessment" provides an immediate safety score based on location and time, complemented by a "Live Crime Feed" that offers real-time alerts from news sources, enhancing a citizen's awareness and ability to make informed safety decisions.

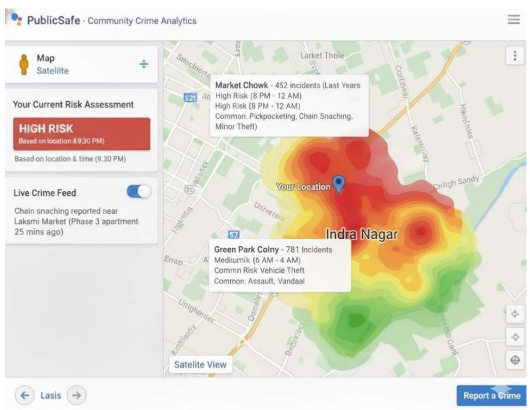

Figure 5 Crime Hotspot Dashboard

V. Scope Representation

Figure 6 Predicted Crime Risk Over 24 Hours

Figure.6 This Predicted Crime Risk graph for Indra Nagar shows that the risk level fluctuates significantly over a 24-hour period. Starting low from midnight to 6 AM, the risk escalates sharply in the afternoon and evening. The data indicates that the highest risk is expected between 3 PM and 9 PM, while the lowest risk is expected from 12 AM to 3 AM. The current time is marked at 9 PM as a period of HIGH RISK.

VI. Conclusion

The dashboard establishes a strong foundation for bridging the gap between complex crime data and public safety awareness. By combining a predictive XGBoost model, simulated real-time data from NewsAPI, and a user-friendly Streamlit interface, it successfully reduces the complexity of personal risk assessment. Beyond just predicting hotspots, the platform contributes to community engagement by offering transparent historical trend analysis, ensuring that the technology augments human judgment with accessible, data-driven insights.

Furthermore, the platform shows promise in reducing the uncertainty and stress associated with navigating urban environments while improving access to safety intelligence for all. With its scalability and adaptability across different cities, it represents a robust model for future digital public safety ecosystems within smart city initiatives.

Acknowledgment

The authors express their sincere gratitude to Bangalore Institute of Technology, Bengaluru, and the Department of Information Science and Engineering for providing guidance, resources, and continuous support throughout this work. Special thanks are extended to Dr. Shilpa M., Associate Professor, for her valuable mentorship and direction. The team also acknowledges the encouragement of faculty members, peers, and collaborators whose contributions have been instrumental in completing this research successfully.

References

[1] N. Kanimozhi, N. V. Keerthana, G. S. Pavithra, G. Ranjitha and S. Yuvarani, "CRIME Type and Occurrence Prediction Using Machine Learning Algorithm," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 266-273, doi: 10.1109/ICAIS50930.2021.9395953.

[2] K. T. M, L. T. N, M. Ithihas, N. R. Shetty, A. H. N and S. Hebbar, "Crime Type and Occurrence Prediction Using Machine Learning," 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 2024, pp. 1-5, doi: 10.1109/ICAIT61638.2024.10690652.

[3] X. Tong et al., "Urban Crime Trends Analysis and Occurrence Possibility Prediction based on Light Gradient Boosting Machine," 2021 IEEE 4th International Conference on Big Data and Artificial Intelligence (BDAI), Qingdao, China, 2021, pp. 98-103, doi: 10.1109/BDAI52447.2021.9515252.

[4] V. Mandalapu, L. Elluri, P. Vyas and N. Roy, "Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions," in IEEE Access, vol. 11, pp. 60153-60170, 2023, doi: 10.1109/ACCESS.2023.3286344.

[5] J. K. Chauhan, A. K. Pandey, A. K. Verma and A. K. Gupta, "Machine Learning Algorithm for Predicting Crime Type and Occurrence," 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI), Greater Noida, India, 2025, pp. 322-327, doi: 10.1109/ICCSAI64074.2025.11064458.

[6] Jangra, Mrinalini, and Shaveta Kalsi. "Naïve Bayes approach for the crime prediction in data mining." International Journal of Computer Applications 178.4 (2019): 33-37.

[7] Ivan, Niyonzima, et al. "Crime Prediction Using Decision Tree (J48) Classification Algorithm." (2017). Kim, Suhong, et al. "Crime analysis through machine learning." 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2018.

[8] Kalyani, G., MVP Chandra Sekhara Rao, and B. Janakiramaiah. "Privacy-preserving classification rule mining for balancing data utility and knowledge privacy using adapted binary firefly algorithm." Arabian Journal for Science and Engineering 43 (2018): 3903-3925.

[9] [5] Almanie, Tahani, Rsha Mirza, and Elizabeth Lor. "Crime prediction based on crime types and using spatial and temporal criminal hotspots." arXiv preprint arXiv:1508.02050 (2015).

[10] Sathyadevan, Shiju, M. S. Devan, and S. Surya Gangadharan. "Crime analysis and prediction using data mining." 2014 First international conference on networks & soft computing (ICNSC2014). IEEE, 2014.

[11] Yuki, Jesia Quader, et al. "Predicting crime using time and location data." Proceedings of the 7th International Conference on Computer and Communications Management. 2019.

[12] Chen, Peng, and Justin Kurland. "Time, place, and modus operandi: a simple apriori algorithm experiment for crime pattern detection." 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA). IEEE, 2018.

[13] Agarwal, Jyoti, Renuka Nagpal, and Rajni Sehgal. "Crime analysis using k-means clustering." International Journal of Computer Applications 83.4 (2013).
[14] Sheng Li, Lihua Yang, Peng Zhang, and Lei Wang. "Crime prediction based on spatiotemporal data using LSTM neural network." Neurocomputing 441 (2021): 251-263. Kai Zhang et al. "A hybrid model based on neural network and decision tree for crime prediction." Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 2, pp. 1553-1562, 2021.
[15] Kai Zhang et al. "A hybrid model based on neural network and decision tree for crime prediction." Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 2, pp. 1553-1562, 2021.
[16] Jiang, Y., Xu, C., Chen, Y., Chen, J., & Chen, H. (2021). DeepCrime: Crime Prediction Based on Deep Learning in Smart City. IEEE Transactions on Industrial Informatics, 17(4), 2485-2494.
[17] Hossain, M. S., Rahman, M. M., & Hossain, M. A. (2021). Predicting the crime hotspots using machine learning techniques: A case study in Dhaka city, Bangladesh. PloS one, 16(4), e0249774.
[18] Sinaga, Kristina P., and Miin-Shen Yang. "Unsupervised K-means clustering algorithm." IEEE access 8 (2020): 80716-80727.
[19] Suthaharan, Shan, and Shan Suthaharan. "Support vector machine." Machine learning models and algorithms for big data classification: thinking with examples for effective learning (2016): 207-235.
[20] Raza, Dewan Mamun, and Debasish Bhattacharjee Victor. "Data mining and Region Prediction Based on Crime Using Random Forest." 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). IEEE, 2021.
[21] Hajela, Gaurav, Meenu Chawla, and Akhtar Rasool. "A clustering based hotspot identification approach for crime prediction." Procedia Computer Science 167 (2020): 1462-1470.