



OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

Crime Pattern Detection Using Data Mining And Machine Learning

Abhay Gaidhani, Harshal Annasaheb Bramhane, Ajay Vishnu Wagh, Darshan Sudhir Revagade, Jay Arun Mahajan

Professor, Computer Engineering, Sandip Institute of Technology and Research Centre

Student, Computer Engineering, Sandip Institute of Technology and Research Centre

abhay.gaidhani@sitrc.org, harshbramhane.tech@gmail.com, awagh1935@gmail.com, darshanrevagade2005@gmail.com

jaymahajan1234.jm@gmail.com

Abstract: Urban safety agencies increasingly require data-driven systems capable of anticipating and localizing criminal activity before it occurs. *Crime Analytics Pro* presents an integrated framework that combines machine-learning-based prediction, time-series forecasting, and geospatial hotspot visualization to support proactive decision-making. The platform automates preprocessing of multi-source crime datasets, extracts temporal and spatial features, and employs ensemble and regression models such as Random Forest, Linear Regression, and clustering techniques (K-Means, DBSCAN) to model crime patterns across regions and time horizons. Forecast outcomes are visualized through Folium-based interactive maps and an intuitive Streamlit dashboard designed for non-technical users. The system emphasizes interpretability through feature-importance evaluation, correlation analysis, and forecast confidence intervals while maintaining modular scalability. By merging predictive analytics with geospatial intelligence in a unified environment, *Crime Analytics Pro* enables a comprehensive, transparent, and replicable approach to crime prevention and smart-city governance.

Keywords: Crime Prediction; Machine Learning; Random Forest; Time Series Forecasting; Hotspot Mapping; Geospatial Intelligence; Folium; Streamlit Dashboard; Public Safety Analytics; K-Means; DBSCAN; Urban Crime Forecasting; Interpretability; Cross-Validation

I. INTRODUCTION

The rapid urbanization of modern societies has led to a significant rise in the complexity and volume of public safety challenges. Law enforcement agencies and city administrators increasingly depend on analytical tools to understand the underlying dynamics of criminal activity and to develop preventive strategies. Traditional crime analysis methods—primarily reliant on manual record inspection and static statistical summaries—often fail to capture the nonlinear, spatio-temporal patterns that characterize real-world criminal behavior. Consequently, the transition from descriptive analytics to predictive and prescriptive intelligence has become a central focus in data-driven policing and urban governance.

Machine Learning (ML) and Data Science offer powerful mechanisms to identify hidden correlations and evolving trends in large and heterogeneous crime datasets. Algorithms such as Random Forest, Linear Regression, and clustering models can learn from historical data to infer future trends and geographical hotspots. However, despite the rapid progress in computational intelligence, most existing systems remain fragmented or domain-specific, lacking an integrated and interpretable framework that combines predictive modeling, time-series forecasting, and

geospatial visualization within a single, user-friendly environment. The absence of such integration limits the operational value of analytical insights for decision-makers and field officers.

Crime Analytics Pro addresses these limitations through a unified architecture that integrates machine learning-based crime prediction with geospatial intelligence and time-series analysis. The system automates the preprocessing of raw crime data, performs spatial and temporal feature engineering, and provides intuitive interactive dashboards for visualization and exploration. This approach facilitates not only predictive forecasting but also interpretability and transparency, enabling end-users to understand how input features influence model outputs. Moreover, the modular design of the platform ensures scalability and extensibility for future incorporation of advanced algorithms and larger datasets.

By bridging the gap between predictive analytics and operational usability, *Crime Analytics Pro* supports proactive crime prevention, strategic resource allocation, and evidence-based urban planning. The framework thus represents a step toward transparent, data-driven public safety systems that combine technological innovation with practical decision support for governance and law enforcement.

II.LITERATURE SURVEY

1. **Crime prediction through machine learning.**
Recent studies highlight that ensemble learning algorithms such as **Random Forest** and **Gradient Boosting** consistently outperform traditional regression-based models in capturing nonlinear crime patterns and variable interactions [1]. These methods enhance classification accuracy and provide interpretable feature-importance metrics that guide law enforcement planning [2].
2. **Spatio-temporal analysis of criminal activity.**
Research has established that combining spatial and temporal variables yields stronger predictive power than either dimension alone [3]. Integrating geospatial encodings (latitude, longitude, neighborhood identifiers) with time-series components (hour, weekday, season) allows models to uncover cyclical or hotspot-based trends in urban environments [4].
3. **Use of clustering algorithms for hotspot identification.**
Unsupervised algorithms such as **K-Means** and **DBSCAN** have been effectively used to detect emerging crime clusters and high-density regions [5]. DBSCAN's ability to handle irregularly shaped clusters and outliers makes it particularly suited for heterogeneous urban data [6].
4. **Time-series forecasting in crime analytics.**
Seasonal ARIMA, exponential smoothing, and more recent neural architectures have been explored for short-term and long-term crime forecasting [7]. Incorporating trend and seasonality decomposition improves forecast interpretability and operational utility for patrol scheduling and policy formulation [8].
5. **Integration of geospatial intelligence and visualization.**
GIS-based visualization tools such as **Folium** and **GeoPandas** enable interactive heatmaps and spatial overlays that help stakeholders intuitively interpret data patterns [9]. Studies emphasize that such visualization bridges the gap between data scientists and non-technical users, improving situational awareness [10].
6. **User-centric dashboards and decision support systems.**
Web frameworks like **Streamlit** and **Dash** have been adopted to develop interactive dashboards for crime analytics [11]. These platforms facilitate real-time data exploration, model retraining, and scenario analysis without requiring specialized programming knowledge [12].
7. **Interpretability and ethical considerations.**
Recent literature underlines the importance of transparent ML models in sensitive domains such as public safety [13]. Explainable AI techniques, including

feature-importance ranking, partial-dependence visualization, and fairness auditing, are critical for trust and accountability in predictive policing applications [14].

8. **Unified frameworks for predictive policing.**
A few modern systems attempt to combine machine learning, geospatial analytics, and visualization, but they often lack modularity or real-time adaptability [15]. *Crime Analytics Pro* extends this line of research by emphasizing scalability, interpretability, and end-user accessibility in a single platform [16].

III.PROBLEM STATEMENT

The existing crime analysis systems are primarily descriptive and reactive in nature. They provide limited capability for predictive modeling and lack the integration of spatial and temporal analytics necessary for proactive decision-making. Most available platforms focus on isolated functionalities such as crime mapping, statistical reporting, or model training without combining them into a unified, user-friendly interface. Moreover, current solutions often fail to ensure interpretability, scalability, and operational usability for non-technical users such as law enforcement officers or policy-makers. There is, therefore, a strong need for an integrated, interpretable, and scalable crime analytics framework that merges machine learning, time-series forecasting, and geospatial intelligence to support proactive crime prediction and informed governance.

IV.OBJECTIVE

General Objective

To develop an integrated and intelligent platform that utilizes machine learning, time-series forecasting, and geospatial analytics for accurate crime prediction, hotspot identification, and decision support in public safety management.

Specific Objectives

1. To preprocess and clean multi-dimensional crime datasets for improved data quality and analytical accuracy.
2. To design and implement predictive models using algorithms such as Random Forest, Linear Regression, and clustering techniques (K-Means and DBSCAN) for spatio-temporal crime prediction.
3. To perform time-series forecasting to identify periodic trends and future crime patterns.
4. To generate interactive heatmaps and geospatial visualizations using Folium for hotspot detection.
5. To build an intuitive, web-based dashboard using Streamlit for end-user interaction, data upload, model selection, visualization, and report generation.
6. To enhance interpretability through correlation analysis, feature importance metrics, and confidence interval visualizations.

7. To ensure modularity, scalability, and low-latency performance for real-time analysis and decision support.



Figure 1 – System Architecture

V.SYSTEM ARCHITECTURE

The proposed system architecture of *Crime Analytics Pro* represents an end-to-end AI-enabled geospatial forecasting and visualization framework designed for predictive crime analytics. It integrates data ingestion, preprocessing, machine learning, geospatial intelligence, evaluation, and visualization within a modular and scalable environment.

1. Actors and Data Sources

- **Law Enforcement / Analyst:** Primary user responsible for uploading datasets, triggering training or inference, and interpreting outputs for decision support.
- **Crime Data Sources:** Structured or semi-structured inputs (CSV, JSON, SQL) containing incident-level details such as crime type, time, and location.
- **Contextual Data Feeds:** Auxiliary information such as demographics, weather, or mobility data used for feature enrichment.

2. Processing Core

The central *Crime Analytics Pro* engine performs the following sequential operations:

- **Data Ingestion & Integration:** Validates file formats, enforces schema consistency, and captures metadata for provenance.
- **Preprocessing Module:** Cleans and transforms data through imputation, outlier handling, temporal binning, and spatial encoding.
- **Machine Learning Engine:** Trains and deploys models for regression, classification, and time-series forecasting (e.g., Random Forest, XGBoost, ARIMA/Prophet) with cross-validation support.
- **Geospatial Analysis:** Conducts spatial joins, clustering, and hotspot detection (KDE, Getis-Ord Gi*) to identify

high-risk regions.

3. Downstream Services

- **Model Evaluation:** Computes quantitative metrics (RMSE, MAE, precision, recall) and spatial validation indicators.
- **Storage Layer:** Persists cleaned datasets, model artifacts, metrics, and prediction outputs in relational or spatial databases(SQLite/SpatiaLite).
- **Visualization Dashboard:** Provides an interactive Streamlit-based interface with heatmaps, trend charts, and downloadable reports.

4. Data Flow Summary

Uploaded data is validated and preprocessed, then passed to the ML engine for predictive modeling. Generated predictions are spatially processed by the GIS layer, stored with evaluation results, and visualized through the web dashboard for operational insights.

Expected Outcome / Results

The proposed *Crime Analytics Pro* framework is expected to deliver an integrated, intelligent environment for predictive crime analysis and visualization. The system combines analytical precision with operational usability, providing actionable insights for urban safety management and policy formulation.

At the functional level, the platform will enable automated ingestion and preprocessing of heterogeneous crime datasets, thereby ensuring data quality and consistency across multiple sources. The machine learning module will generate predictive models capable of recognizing complex spatio-temporal patterns, facilitating proactive identification of emerging hotspots and potential risk zones. The integration of time-series forecasting will allow agencies to anticipate crime trends over varying temporal horizons, enhancing situational preparedness and resource allocation.

The geospatial analysis component will support interactive mapping and hotspot visualization through Folium-based interfaces, offering law enforcement agencies a clear spatial understanding of risk concentrations. The visualization dashboard will further enhance decision-making by enabling real-time exploration of data, model outputs, and derived insights through intuitive visual analytics.

From a non-functional perspective, the architecture is designed for modularity, scalability, and interpretability. It ensures transparent model behavior through feature importance visualization and correlation analysis, while the storage and evaluation layers maintain auditability and reproducibility. Collectively, these outcomes aim to establish *Crime Analytics Pro* as a robust, extensible, and ethically aligned solution for smart, data-driven governance in public safety domains.

VI.REFERENCES

- [1] S. K. Gupta and A. Kumar, "Machine learning-based crime prediction using ensemble classifiers," *International Journal of Data Science and Analytics*, vol. 12, no. 3, pp. 245–258, 2023.

- [2] L. Zhang, J. Chen, and M. Wang, "Comparative study of Random Forest and Gradient Boosting for urban crime forecasting," *Applied Intelligence*, vol. 53, no. 5, pp. 4201–4216, 2024.
- [3] H. Patel and P. Shah, "Spatio-temporal analysis of criminal incidents using predictive analytics," *Procedia Computer Science*, vol. 225, pp. 151–160, 2023.
- [4] J. R. Silva, M. Brito, and C. Almeida, "Time-aware spatial modeling of crime trends in metropolitan regions," *Journal of Urban Computing*, vol. 11, no. 2, pp. 110–124, 2023.
- [5] D. Singh and R. Kaur, "Crime hotspot detection using K-Means and DBSCAN clustering algorithms," *Pattern Recognition Letters*, vol. 175, pp. 94–102, 2024.
- [6] K. Liu and F. Li, "Density-based spatial clustering for irregular hotspot detection in heterogeneous urban data," *Information Systems Frontiers*, vol. 26, no. 1, pp. 233–247, 2024.
- [7] A. N. Mehta and R. Patidar, "Forecasting of criminal activities using time-series models," *Expert Systems with Applications*, vol. 226, 102960, 2023.
- [8] P. Verma, R. Ghosh, and S. Saxena, "Trend decomposition and seasonality modeling for predictive policing," *IEEE Access*, vol. 12, pp. 78541–78555, 2024.
- [9] C. Yang, H. Wu, and J. Cheng, "Geospatial data visualization and mapping for crime pattern analysis," *Computers, Environment and Urban Systems*, vol. 106, 102919, 2024.
- [10] R. Torres and B. Martinez, "Interactive GIS for public safety analytics: An open-source implementation," *ISPRS International Journal of Geo-Information*, vol. 13, no. 2, 2024.
- [11] D. Johnson and E. Thomas, "Interactive dashboards for data-driven policing using Streamlit and Dash," *Software: Practice and Experience*, vol. 54, no. 1, pp. 33–49, 2024.
- [12] S. Agarwal and M. Yadav, "Design of intelligent visualization frameworks for urban security analytics," *International Journal of Computer Applications*, vol. 186, no. 10, pp. 15–21, 2023.
- [13] L. Ribeiro and G. Santos, "Explainable artificial intelligence for ethical predictive policing," *AI & Society*, vol. 39, pp. 881–895, 2024.
- [14] E. Howard, T. Brown, and K. Lee, "Fairness-aware machine learning in public safety decision systems," *IEEE Transactions on Technology and Society*, vol. 5, no. 2, pp. 150–164, 2024.
- [15] J. Kim, S. Lee, and D. Park, "Integrated crime analytics using hybrid spatio-temporal models," *IEEE Access*, vol. 11, pp. 121960–121972, 2023.
- [16] H. Bramhane, "Crime Analytics Pro: An integrated AI and geospatial framework for predictive crime analysis," *Academic Project Report*, Himalayan University, 2025.