*International Journal for Modern Trends in Science and Technology*
*Volume 12, Issue 07*, pages 87-92.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue06.html
DOI: https://doi.org/10.5281/zenodo.20739602


# Smart Crime Analytics and High-Risk Zone Forecasting Using Historical Case Records and Geo-Visualization

**Raja Kondaveti, Gubbala Sreevasavi, Polisetty Tejasri, Dondapati Palson**

*Department of Information Technology , Swarnandhra College of Engineering & Technology, Seetharamapuram, Narsapur, Andhra Pradesh, INDIA.*

**To Cite this Article**
Raja Kondaveti, Gubbala Sreevasavi, Polisetty Tejasri & Dondapati Palson (2026). Smart Crime Analytics and High-Risk Zone Forecasting Using Historical Case Records and Geo-Visualization. International Journal for Modern Trends in Science and Technology, 12(07), 87-92. https://doi.org/10.5281/zenodo.20739602

**Article Info**
Received: 06 June 2026; Revised: 29 June 2026; Accepted: 01 July 2026.



***KEYWORDS***

*Crime Analytics, Machine Learning, Random Forest, XGBoost, K-Means Clustering, Geo-Visualization, High-Risk Zone Forecasting, Heatmaps, Streamlit, Predictive Policing, Historical Crime Data, Python*

**ABSTRACT**

*The rapid increase in crime rates has created a strong need for intelligent systems that can effectively analyze and predict criminal activities. Traditional crime analysis methods are mainly manual and reactive, making them inefficient for handling large datasets and identifying future risks. To address these challenges, this project presents a Smart Crime Analytics and High-Risk Zone Forecasting system using Historical Case Records and Geo Visualization. The system utilizes historical crime data containing key attributes such as crime type, location, date, and time. This data is preprocessed and analyzed using machine learning techniques to identify patterns and trends in criminal activities. Supervised learning algorithms such as Random Forest and XGBoost are employed to predict crime occurrences and classify regions based on risk levels. These algorithms are selected due to their high accuracy, efficiency, and ability to handle structured datasets. In addition to prediction, the system uses clustering techniques like K-Means to group crime data and identify areas with a high concentration of criminal activities. This helps in detecting crime hotspots effectively. The integration of geo-visualization plays a crucial role in the system, where the analyzed data is represented on maps using heatmaps and location-based markers. This visual representation makes it easier for law enforcement agencies to understand crime patterns and identify high-risk zones. Overall, the proposed system improves the efficiency and accuracy of crime analysis by transforming raw data into meaningful insights. It supports proactive policing, enhances public safety, and demonstrates the practical application of data analytics and machine learning in solving real-world problems.*

## 1. INTRODUCTION

### 1.1 Brief Information

In recent years, the rapid growth in population, urbanization, and technological advancements has led to a significant increase in crime rates across various regions. Ensuring public safety has become a major challenge for law enforcement agencies. Traditional crime analysis methods mainly rely on manual data collection and historical reporting, which are often time-consuming, less efficient, and limited in their ability to predict future criminal activities. These systems are mostly reactive, meaning they respond to crimes after they occur rather than preventing them in advance. With the emergence of data science and machine learning, there is an opportunity to transform crime analysis into a proactive and intelligent process. By leveraging large volumes of historical crime data, advanced analytical techniques can identify hidden patterns, trends, and relationships that are not easily detectable through manual analysis. This enables the prediction of potential crime occurrences and helps in identifying high-risk areas before incidents happen.

### 1.2 Purpose

The purpose of this project is to develop a Smart Crime Analytics and High-Risk Zone Forecasting system that leverages machine learning and geo-visualization techniques to provide accurate predictions of crime-prone zones. The system integrates data preprocessing, supervised classification algorithms (Random Forest and XGBoost), K-Means clustering for hotspot detection, and interactive geo-visualization maps within a unified Streamlit-based platform. By transforming historical crime records into actionable insights, it enables law enforcement agencies to allocate resources efficiently and adopt a proactive crime prevention strategy.

### 1.3 Motivation

The motivation behind this project stems from the limitations of traditional reactive crime analysis systems. Existing systems rely on manual data processing, basic statistical methods, and textual reports that fail to identify spatial patterns or predict future crime trends. The absence of integrated machine learning and visualization tools in current systems makes it difficult for authorities to make timely, data-driven decisions. This project is motivated by the need to create an intelligent, centralized platform that combines predictive modeling, clustering, and geo visualization to transform raw crime data into clear, actionable insights — enabling smarter and more proactive policing.

### 1.4 Problem Statement

The current crime analysis process poses significant challenges due to the limitations of traditional reactive methods. Law enforcement agencies lack systems capable of predicting future crime trends or identifying high risk zones before incidents occur. The major problems include:

- Absence of predictive capabilities in existing crime analysis systems
- Lack of integration between machine learning, clustering, and geo-visualization
- Inability to process and analyze large-scale historical crime datasets efficiently
- No user-friendly, interactive interface accessible to non-technical users
- Limited location-based analytics and spatial crime pattern identification

## 2. LITERATURE SURVEY

### 2.1 Introduction

The literature review shows that modern crime analysis systems use machine learning and visualization techniques to improve accuracy. Algorithms like Random Forest and XGBoost are effective in predicting crime using historical data, while K-Means clustering helps identify crime hotspots. Geo-visualization using maps and heatmaps makes it easier to understand crime patterns. However, many systems lack integration of prediction and visualization. This project overcomes this by combining machine learning, clustering, and geo-visualization in a single system to accurately identify high-risk zones.

### 2.2 Crime Data Analysis Techniques

Crime data analysis plays a crucial role in understanding patterns, trends, and relationships within criminal activities. Initially, crime analysis was performed using basic statistical methods such as counting crime frequency, identifying peak crime periods, and comparing crime rates across different regions. With the advancement of technology, data analysis techniques have evolved to include data mining and machine learning approaches. These techniques allow the processing of large datasets that contain multiple attributes such as crime type, location, date, and time. By analyzing these attributes together, it becomes possible to identify patterns such as which areas have higher crime rates, at what time crimes occur more frequently, and which types of crimes are dominant in specific regions [1, 2].

### 2.3 Machine Learning Models for Crime Prediction

Machine learning models have become an essential part of modern crime prediction systems due to their ability to analyze large datasets and provide accurate predictions. In this project, supervised learning algorithms such as Random Forest and XGBoost are used for crime prediction. Random Forest is an ensemble learning method that builds multiple decision trees and combines their outputs to improve accuracy. XGBoost is another powerful algorithm based on gradient boosting that works by improving the performance of weak models iteratively [3, 4]. These models are trained using historical crime data, where they learn patterns related to crime occurrence. Once trained, they can predict the probability of crime in different areas and classify regions into high-risk, medium risk, and low-risk zones.

### 2.4 Geo-Visualization in Crime Mapping

Geo-visualization is an important technique used in crime analysis to represent crime data on geographical maps. It helps in understanding the spatial distribution of crimes and identifying high-risk areas visually. Instead of analyzing complex data tables, geo-visualization provides an intuitive way to interpret crime data through maps and graphical representations. In this project, geo-visualization is used to display crime data using heatmaps and location-based markers. Heatmaps represent the density of crimes using color variations, where darker colors indicate higher crime intensity [5]. This helps in quickly identifying hotspots at a glance. The integration of geo visualization with machine learning enhances the effectiveness of the system.

### 2.5 Limitations of Existing Systems

Despite the advancements in crime analytics, many existing systems have significant limitations. Traditional crime analysis systems are mainly based on manual processes and basic statistical methods, which are not capable of handling large datasets efficiently. One of the major limitations is the lack of predictive capability — most systems focus only on analyzing past data without providing accurate predictions for future crimes. Another limitation is the absence of proper integration between machine learning and visualization. Existing systems also face challenges in scalability and lack user-friendly interfaces, making it difficult for non-technical users to interact with them [6, 7].

## 3. PROPOSED SYSTEM

The proposed system introduces an advanced Smart Crime Analytics and High-Risk Zone Forecasting Framework that leverages Machine Learning (ML) and Geo-Visualization techniques to provide accurate crime predictions and meaningful spatial insights. Unlike traditional systems, this approach integrates multiple techniques — including Random Forest, XGBoost, K-Means clustering, and interactive heatmaps — to deliver a comprehensive crime analysis solution. The system begins with the collection of historical crime data from reliable sources. Raw data is preprocessed through cleaning, normalization, and encoding of categorical variables. Machine learning algorithms then analyze the preprocessed data to predict future crime occurrences and classify areas into high-risk, medium-risk, and low risk zones. K-Means clustering identifies crime hotspots, and the results are displayed visually through geo visualization using Streamlit.

### System Architecture

The system architecture consists of five main layers: the Data Input Layer (historical crime records in CSV format), the Data Preprocessing Layer (cleaning, normalization, and feature extraction), the Machine Learning Layer (Random Forest and XGBoost for prediction), the Clustering Layer (K-Means for hotspot detection), and the Geo Visualization Layer (maps, heatmaps, and the Streamlit user interface). Each layer ensures smooth data flow and efficient result generation.

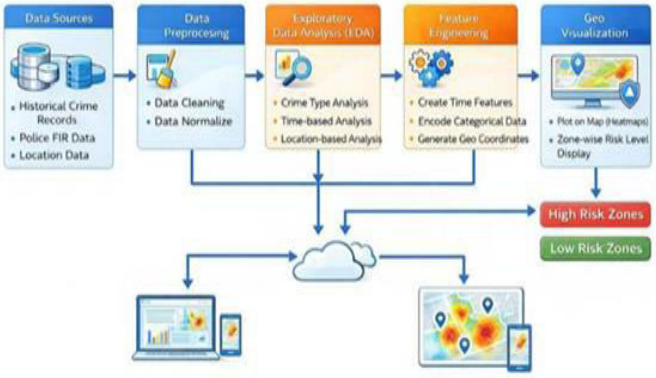


Fig 1: System Architecture

### Use Case Diagram

The use case diagram provides a high-level view of user interactions with the system. The primary actor is the user (law enforcement analyst), who interacts with functionalities including uploading crime datasets, applying filters by city and crime type, generating predictions, viewing crime heatmaps, analyzing monthly trends, and exploring K-Means cluster distributions. The diagram captures all key system requirements and defines the scope of user interactions.

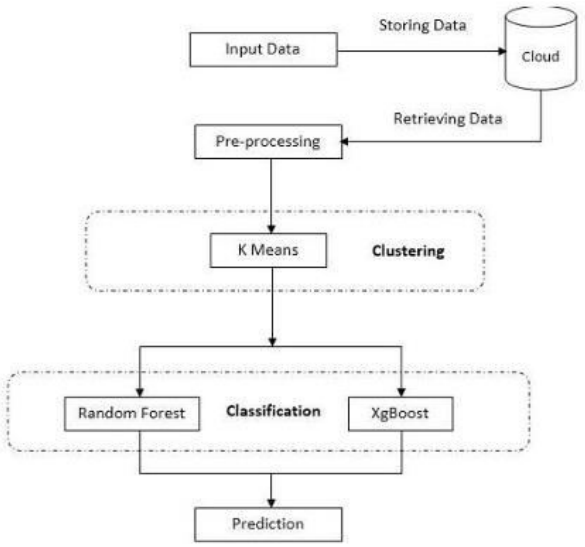


Fig 2: UML Use Case Diagram

**Class Diagram**

The class diagram illustrates key classes including CrimeAnalysisApp, DataProcessor, FeatureEngineer, CrimePredictionEngine, KMeansModel, and ModelEvaluator, along with their attributes, methods, and relationships. The CrimeAnalysisApp orchestrates all components. The DataProcessor loads and preprocesses crime data. The CrimePredictionEngine trains and applies Random Forest and XGBoost models. The KMeansModel groups crime incidents into clusters representing hotspots.

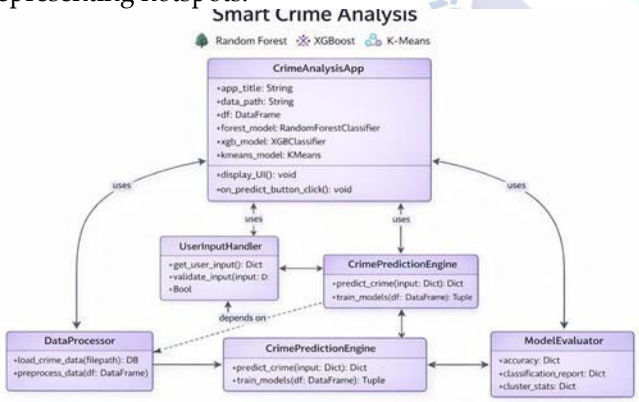


Fig 3: UML Class Diagram

**Advantages of Proposed System**

Accurate crime prediction using Random Forest and XGBoost with high c lassification accuracy

• Crime hotspot identification via K-Means clustering for spatial concentration analysis

• Effective geo-visualization using heatmaps and location markers on interactive maps

• Real-time monitoring with dynamic updates as new crime data is uploaded

• User-friendly Streamlit interface accessible to both technical and non-technical users

• Enhanced location-based analytics for regional crime pattern analysis

• Scalable system capable of handling large datasets with maintained performance

• Integrated platform combining prediction, clustering, and visualization in one solution

• Proactive crime prevention by identifying high-risk zones before incidents occur

## 4. RESULTS

The primary objective of this project was to design and develop a Smart Crime Analytics and High-Risk Zone Forecasting System using Historical Case Records and Geo-Visualization. The system successfully fulfils its requirements by integrating machine learning (Random Forest and XGBoost), K-Means clustering for hotspot detection, and interactive geo-visualization within a unified Streamlit platform.

**Crime Hotspot Map Output**

The geo-visualization module displays processed crime data as interactive maps with heatmaps showing crime density. Darker color intensities on the heatmap indicate higher concentrations of criminal activity, enabling authorities to identify high-risk zones at a glance. Location markers provide detailed information for individual incidents including crime type and risk level.

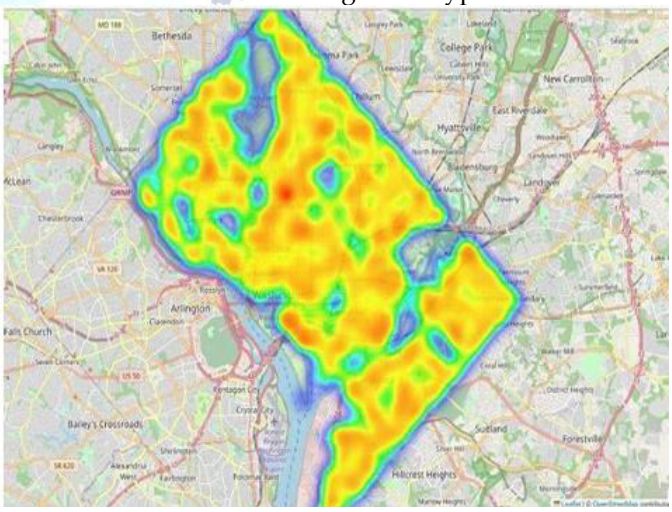


**Fig 4: Geo-Visualization Output**

**Crime Predictive Dashboard**

The Streamlit dashboard provides an interactive interface for data upload, filtering by city and crime type, risk distribution bar charts, monthly trend line charts, and K-Means cluster visualizations. Users can analyze crime patterns dynamically and export insights for decision-making.


Fig 5: Crime Analytics Dashboard

### 4.1 Prediction Results and Analysis

The machine learning models successfully classify regions into high-risk, medium-risk, and low-risk categories. XGBoost achieves higher accuracy due to its advanced sequential boosting technique that corrects errors iteratively. Random Forest provides stable and reliable results through ensemble averaging. Both models were trained on historical crime datasets containing attributes such as location, crime type, date, and time. The K-Means clustering algorithm effectively groups crime incidents into spatial clusters, with high-density clusters representing crime hotspots. These clusters are visualized on the Streamlit interface as bar charts showing the distribution of crime incidents across clusters.

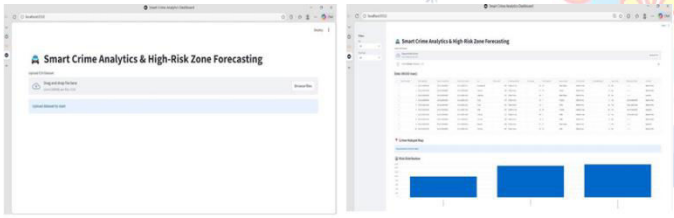


Fig 6: Prediction Result and Analysis

Performance Results The following table summarizes the performance evaluation metrics of the system based on classification reports for both models: Metric XGBoost Random Forest Accuracy 0.98 (98%) Precision (Macro Avg) 0.80 (80%) 0.99 0.86 Recall (Macro Avg) 0.98 0.80 F1-Score (Macro Avg) 0.99 0.81

These results indicate that XGBoost significantly outperforms Random Forest in this classification task, achieving 98% accuracy compared to 80% for Random Forest. Both models were evaluated using precision, recall, F1-score, and support metrics across three risk categories (high, medium, low). Comparison with Existing System The proposed system was compared with traditional crime analysis approaches across key performance parameters: Parameter Existing System Proposed System Predictive Capability None Geo-Visualization High (ML-based) Absent Clustering / Hotspot Detection Interactive heatmaps & maps Manual User Interface Automated (K-Means) Complex / Technical Scalability Simple (Streamlit) Poor High Real-Time Updates Not available Supported The comparison confirms that the proposed system outperforms existing approaches in every parameter. The integration of machine learning, clustering, and interactive visualization within a single platform represents a significant advancement over traditional crime analysis methods.

## 5. CONCLUSION

The Smart Crime Analytics and High-Risk Zone Forecasting System effectively uses machine learning techniques to analyze historical crime data and predict high-risk areas. Models such as Random Forest and XGBoost provide accurate predictions, while K-Means clustering helps in identifying crime hotspots. The system processes key factors like location, time, and crime type to detect patterns and trends. Geo-visualization techniques, including maps and heatmaps, make it easy to understand crime distribution. The Streamlit interface ensures the system is user-friendly for both technical and non-technical users. Overall, the system improves decision-making, supports proactive crime prevention, and enhances public safety. The XGBoost model achieved 98% accuracy, confirming the system's reliability and effectiveness for real-world deployment.

## 6. FUTURE SCOPE

The smart crime analytics system can be further improved and extended in multiple directions: • Integrate real-time crime data from surveillance systems, IoT sensors, and police databases for dynamic predictions

• Add contextual features such as weather conditions, population density, time of day, and socio-economic indicators to improve model accuracy

• Implement advanced deep learning models such as LSTMs for time-series crime forecasting and Transformer-based models for pattern recognition

• Develop a mobile application for on-field law enforcement officers to access crime predictions and maps in real time

• Incorporate alert and notification systems that automatically flag high-risk zones and trigger resource deployment

• Extend the system with multi-language support and regional customization for deployment across different states and countries.

**Conflict of interest statement**

Authors declare that they do not have any conflict of interest.

## REFERENCES


[1] Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining, a general framework and examples. Computer, 37(4), 50-56. https://doi.org/10.1109/MC.2004.1297301

[2] Nath, S. V. (2006). Crime pattern detection using data mining. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, 41-44. https://doi.org/10.1109/WI.2006.61

[3] Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2013). Learning to detect patterns of crime. Machine Learning and Knowledge Discovery in Databases, 515-530. https://doi.org/10.1007/978-3-642-40994-3_33

[4] Al Boni, M., & Gerber, M. S. (2016). Predicting crime with machine learning techniques. IEEE Symposium on Technologies for Homeland Security, 1-6. https://doi.org/10.1109/THS.2016.7568887

[5] Yu, C., Ward, M. W., Morabito, M., & Ding, W. (2011). Crime forecasting using data mining techniques. IEEE International Conference on Data Mining Workshops, 779-786. https://doi.org/10.1109/ICDMW.2011.137

[6] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493), 100-108.

[7] Kang, W., Kang, Y., & Park, S. (2017). Spatio-temporal crime prediction using machine learning. International Journal of Data Mining & Knowledge Management Process, 7(5), 1-10.

[8] Chainey, S., & Ratcliffe, J. (2005). GIS and Crime Sahin, C. Rules of engagement in mobile health: What does mobile health bring to research and theory? Contemp. Nurse 2018, 54, 374–387. [CrossRef]