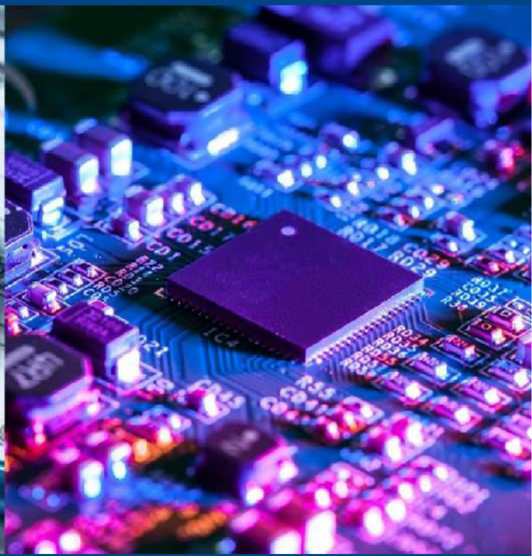
ISSN: 2582-7219

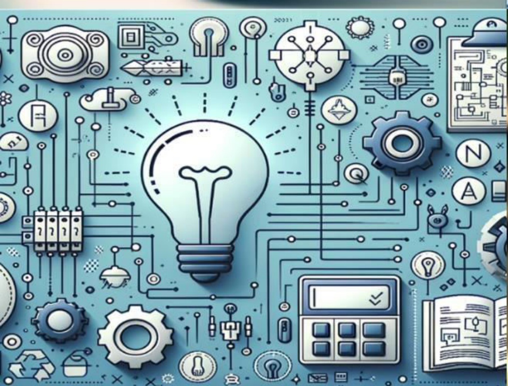




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 10, October 2025

© 2025 IJMRSET | Volume 8, Issue 10, October 2025| DOI:10.15680/IJMRSET.2025.0810038

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Crime Prediction using Machine Learning

Abhishek S, Chandrakant G K, Sandeepgouda A S, Varun V S, Prof. Vani G S

UG Student, Dept. of ISE, PES Institute of Technology and Management, Shivamogga, Karnataka, India

Assistant Professor, Dept. of ISE, PES Institute of Technology and Management, Shivamogga, Karnataka, India

ABSTRACT: Urban crime continues to challenge public safety and community well-being, with conventional policing often limited to reactive measures. This research investigates how machine learning can enable proactive crime prevention by forecasting incidents before they occur, leveraging historical patterns and contextual factors. Using open-source crime datasets, we developed a predictive system that analyzes temporal, spatial, and categorical crime indicators. Multiple algorithms—including Decision Trees, Support Vector Machines (SVM), and Random Forests—were evaluated, with Random Forests achieving superior performance. Beyond predictions, the system incorporates interactive visualization tools such as dynamic heatmaps and Live crime risk maps to display high-risk areas in real time. The results demonstrate the potential of data-driven approaches to augment policing strategies. By integrating predictive analytics, this work contributes to more informed resource allocation and timely interventions, thereby paving the way for safer urban environments.

I. INTRODUCTION

In modern cities—where populations are dense and life moves quickly—ensuring public safety has become increasingly complex. Crime not only endangers individuals but also weakens economic stability and damages community trust. Police and law enforcement agencies dedicate immense effort to combating criminal activity, yet their traditional methods often rely on reactive measures and manual processes, limiting their effectiveness. But what if we could anticipate crimes before they occur? With the growing availability of crime-related data, such as incident locations, timings, and crime types, we can now leverage machine learning to transform this information into valuable insights. This research explores a machine learning- based system designed to analyse historical crime data, detect hidden patterns, and predict future criminal activity. By processing key factors like time, crime category, and geographical data, the system identifies high-risk areas and periods for specific offenses. The study has two primary objectives: first, to evaluate the effectiveness of different machine learning models in predicting crime based on records, and second, to develop an interactive platform that presents these findings in a clear and user-friendly manner. Features such as heatmaps and Live crime risk maps will assist law enforcement in making data-driven decisions while keeping communities informed and engaged. Ultimately, this project highlights how machine learning can contribute to crime prevention, not by replacing human expertise but by empowering authorities with predictive analytics, timely alerts, and proactive strategies to create safer neighbourhoods.

II.LITERATURE REVIEW

In [1, this review analyzes 150+ studies on crime prediction with ML and DL, highlighting models like Random Forest, SVM, CNNs, and LSTMs using spatial-temporal data from datasets such as Chicago, London, and NYC. Reported accuracies reach 99.9%. Challenges include data quality, interpretability, and ethics, while future directions point to reinforcement learning and feature-focused analysis for improved predictive policing.

In [2], this study used ML models like Linear Regression, Random Forest, and KNN to analyze India's crime patterns. KNN effectively identified high-risk areas, with heatmaps visualizing hotspots for better patrol planning. The system was especially useful in predicting property crimes and thefts, offering a proactive policing approach for developing nations.

In [3], using Vancouver Police data, this project showed XGBoost (56% accuracy) outperforming Random Forest. Adding features like holiday calendars and safe zones improved predictions, while data cleaning and tuning tackled class imbalance. Interactive Google Maps visualizations enabled hotspot monitoring, with potential to scale across

© 2025 IJMRSET | Volume 8, Issue 10, October 2025| DOI:10.15680/IJMRSET.2025.0810038

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Canadian cities.

In [4, this approach mined geotagged tweets for real-time crime detection, with XGBoost and SVM giving the best results. Techniques were used to manage noisy data and class imbalance, and the system generated crime density maps for police dashboards. Future work targets integrating more diverse data sources beyond Twitter.

In [5], using a Kaggle dataset, this study combined Decision Trees with ARIMA to forecast temporal crime patterns in U.S. cities. Grid plot visualizations revealed neighborhood assault and theft trends, helping predict both where and when crimes may rise. The approach could support police scheduling systems.

In [6], this KNN-based system achieved 93% accuracy, sending automatic email alerts for crimes like robbery and murder, aligned with legal codes (e.g., IPC 302). Dynamic maps showed crime waves by time of day, showcasing AI's potential to shift policing from reactive to proactive.

In [7, this review explored ML's role in policing, from facial recognition to risk assessment. While LSTMs and CNNs proved effective for forecasting, concerns about algorithmic bias were highlighted. The authors urged explainable AI and strict auditing to balance innovation with civil liberty protections.

In [8], this study applied Naïve Bayes to gender-based violence data, reaching 99.9% accuracy in detecting patterns of rape and dowry deaths across Indian states. Interactive dashboards tracked yearly crime trends, showing how targeted ML can address critical societal issues.

In [9, this project analyzed 2,000 cases, using PCA with Random Forest to reliably classify violent vs. non-violent crimes. Heatmaps highlighted high-risk neighborhoods, helping investigators prioritize cases by predicted severity.

In [10], this study benchmarked five algorithms on national crime records, with Naïve Bayes achieving 98.94% accuracy. It exposed geographic patterns, such as Madhya Pradesh's high rape incidence, and enabled district-level predictions for local prevention. Expanding with newer data was recommended for greater relevance.

III.METHODOLOGY

The process of building a reliable crime prediction system involves several key stages, from gathering and preparing data to training models and evaluating their performance. This section outlines the approach followed in this study to develop and deploy an effective machine learning-based crime analysis solution.

1. Gathering Crime Data
To build an effective and practical crime prediction system, our methodology was structured around five core stages: data collection, data preprocessing, model training, evaluation, and integration into a web-based application. Each of these steps played a critical role in ensuring the final system was not only accurate but also usable and relevant for real-world crime analysis.

Crime data was gathered from open and reliable sources such as Kaggle and official city databases, including the Indian crime Incident Reports. The datasets contained structured information like incident date, time, location (latitude and longitude), crime type, and case status. For this study, emphasis was placed on temporal and spatial attributes, as they are key indicators of crime patterns and trends. This data served as the foundation for model training and pattern analysis in the prediction system.

2. Preparing the Data for Machine Learning
Once the data was collected, the next challenge was to prepare it for machine learning. Raw datasets are rarely clean, and this case was no exception. We performed extensive preprocessing to enhance the quality of the data and ensure the models could learn effectively from it. Missing entries were addressed through imputation or removal, depending on their importance. Time-related fields were broken down into components such as hour of the day, day of the week, and month, enabling the model to recognize crime patterns based on daily and seasonal cycles. Non-numeric fields like crime type or district were transformed into numerical values using encoding techniques. Additionally, we

normalized the dataset to ensure uniform scales across features, which is especially beneficial for algorithms sensitive to magnitude differences.

3. Building and Training Prediction Models
For model training, we explored a variety of machine learning algorithms, including Logistic Regression, Decision Trees, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests. Each model was trained and tested using a split of the preprocessed dataset, and we applied cross-validation techniques to ensure robustness and reduce the risk of overfitting. Among all the algorithms tested, the Random Forest model emerged as the most effective. It delivered consistently high accuracy and showed resilience to noise and outliers in the data. Its ensemble nature—combining multiple decision trees—allowed it to capture complex relationships within the dataset, making it especially suited for crime classification tasks.

4. Evaluating Model Performance
The trained models were assessed using standard metrics such as accuracy, precision, recall, and F1-score. The Random Forest model achieved about 80% accuracy with high recall, crucial for minimizing missed crime hotspots. Evaluation tools like confusion matrices and classification reports were used to visualize and analyze model performance.

5. Integrating the Model into a Web Application
Finally, the trained model was integrated into a web application that translates technical predictions into accessible, real-time insights for users. This application serves as an interactive platform with three major features: a prediction tool powered by the ML model, a live map that highlights current or predicted crime-prone areas, and a heatmap that visualizes the concentration of different crime types over time and geography. These features were carefully designed to present information in an intuitive and actionable format, enabling law enforcement and the general public to understand and act on crime trends more effectively.

IV. SYSTEM ARCHITECTURE

The proposed Crime Prediction and Analysis System is designed to forecast crime occurrences and visualize spatial-temporal patterns using machine learning and web technologies. The overall architecture follows a modular and layered structure consisting of five main components: data collection, data preprocessing, model training, backend integration, and frontend visualization.

1. Data Collection Layer
The system begins with the Data Collection Layer, which gathers historical crime datasets from publicly available sources such as Kaggle. The dataset includes essential attributes like date, time, location, and crime type, while irrelevant fields are eliminated to maintain focus on key predictive features. This refined dataset serves as the foundation for training and analysis.

2. Data Preprocessing Layer
The Data Preprocessing Layer ensures the data is clean, consistent, and ready for modeling. Operations such as removal of null values, elimination of duplicates, and feature selection are performed to improve data quality. Temporal information is extracted from the date and time columns, and categorical features such as location and crime type are converted into numeric form using label encoding. Normalization is then applied to bring all attributes to a common scale, allowing efficient learning by machine learning algorithms.

3. Model Training Layer
In the Model Training Layer, multiple algorithms—including Logistic Regression, Decision Tree, Support Vector Machine, k-Nearest Neighbors, and Random Forest—are evaluated to determine the most suitable model for crime prediction. Among these, the Random Forest Classifier achieved the best performance with an accuracy of around 80%. This ensemble-based model, consisting of multiple decision trees, effectively reduces overfitting and improves prediction reliability. The trained model is serialized and stored for integration with the web application.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4. Backend Integration Layer

The Backend Integration Layer is implemented using the Flask web framework, which acts as a bridge between the machine learning model and the user interface. It handles user requests, loads the trained Random Forest model, and generates crime predictions based on user-provided inputs such as time and location. Flask communicates with the frontend through RESTful APIs and returns prediction results in JSON format, ensuring smooth real-time data flow between components.

5. Frontend Visualization Layer

The Frontend Visualization Layer is built using ReactJS and provides an interactive web interface for users and law enforcement authorities. It includes three key modules: a crime prediction tool that predicts the most probable crime type for a given location and time; a live crime risk map that visualizes high-risk areas in real time using Leaflet.js with color-coded risk indicators; and a crime heatmap that displays crime intensity by type and month to help identify historical hotspots.

6. System Workflow

The system operates through a seamless workflow: users enter the location and time on the web interface, the data is sent to the backend via API, the trained Random Forest model predicts the potential crime type, and the results are returned to the frontend for visualization on maps and charts. The architecture ensures scalability, real-time interactivity, high prediction accuracy, and an intuitive user experience.

Overall, this architecture enables effective crime forecasting and visualization by combining the strengths of machine learning, data visualization, and web-based interactivity, ultimately supporting data-driven decision-making for public safety and urban planning.

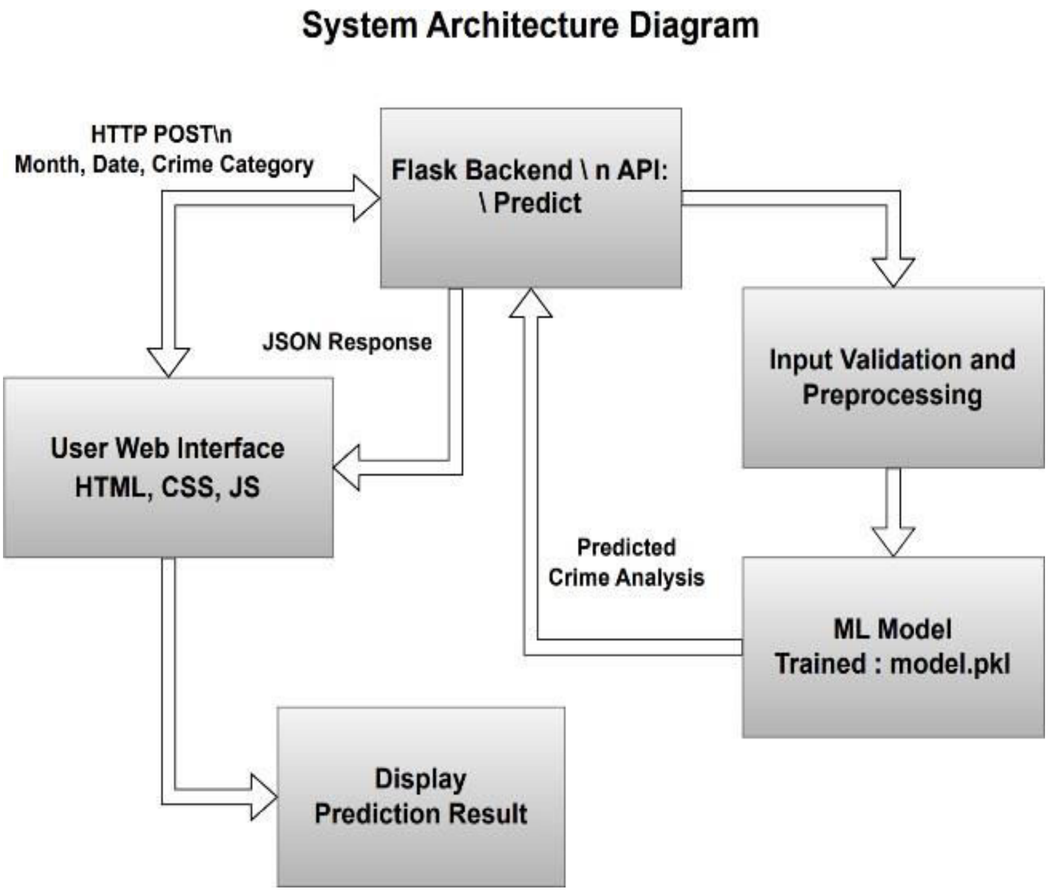


Fig 4.1 System Architecture Diagram.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. RESULTS

1. Crime Prediction Model

The crime prediction model developed using the Random Forest algorithm demonstrated strong performance, achieving an accuracy of 80%. By analyzing time and location attributes from historical crime data, the model was able to predict the most probable type of crime for given conditions. The results were visualized using a probability-based bar graph, which clearly displayed the likelihood distribution across various crime categories, providing an intuitive understanding of prediction outcomes.

Fig 5.1 Crime Prediction results in a bar graph

2. Crime Heatmap Analysis

The Crime Heatmap Analysis offered deeper insight into long-term trends and spatial crime distributions. By visualizing data across different months and crime types, the heatmap effectively identified persistent hotspots and seasonal variations in criminal behaviour. This analysis not only enhanced understanding of crime density patterns but also provided valuable guidance for strategic planning and preventive measures.

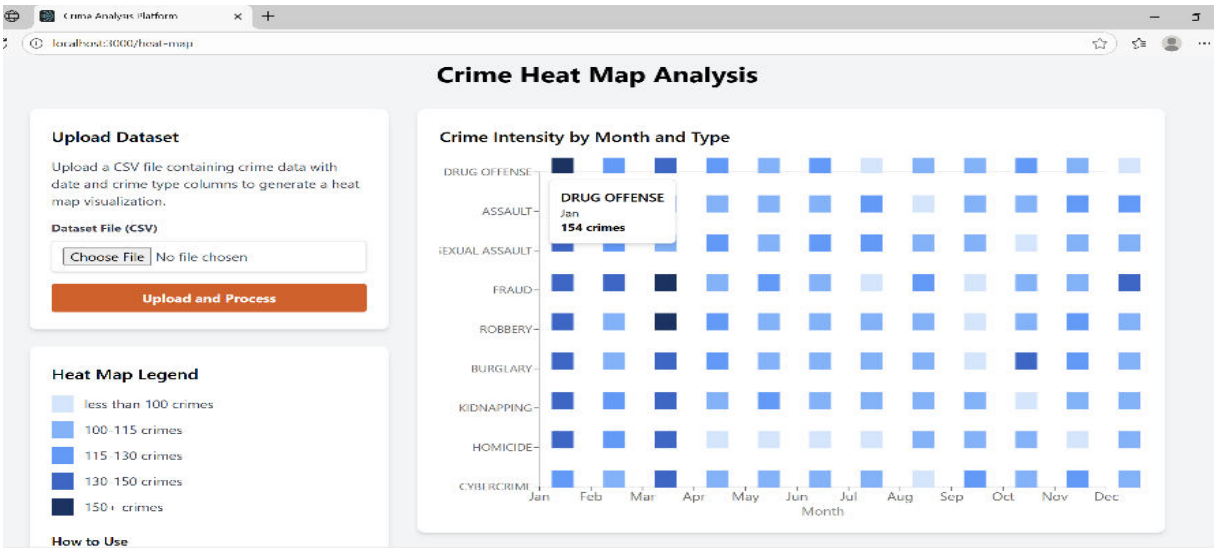


Fig 5.2 Example of Crime analysis using Heatmap

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3. Live Map: Visualizing Crime Risk Zones

The Live Crime Risk Map was designed to represent real-time high-risk areas within a city. It used the trained model's predictions to dynamically highlight zones prone to criminal activity, offering immediate situational awareness. This feature proved useful for both the general public and law enforcement, helping them make informed, location-based safety and deployment decisions.

VI. CONCLUSION

This study demonstrates how machine learning can be effectively used to predict and visualize criminal activity based on historical data. By combining a trained Random Forest model with an interactive web application, we were able to deliver a practical system that not only forecasts likely crimes based on time and location but also helps users understand spatial trends through features like Live Map and Heatmap. The system's ability to provide real-time predictions and compiled crime insights makes it a valuable tool for both public awareness and smarter policing.

REFERENCES

1.Varun Mandalapu, Lavanya Elluri, Piyush Vyas, Nirmalya Roy. "Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions" 2024

2.Sridharan S, Srish N, Vigneswaran S, and Santhi P. " Crime Prediction using Machine Learning" 2024.

3.Sankalp Singh, Akshay Gole, Prathmesh Kanherkar, P R Abhishek, Pallavi Wankhede." Crime Prediction using Machine Learning" 2022.

4.Shivani Bhardwaj, Sandeep Kaur. " Design and Implementation of Crime Prediction Using Machine Learning Algorithm"2022.

5.Mrs. Meena M, Sriabirami V, Divyaa Sri A R. " Analysis and Prediction of Crime Using Machine Learning" 2022.

6.Kavya T M, Lavanya T N, Amrutha H N, "Crime Type and Occurrence Prediction Using Machine Learning" 2024.

7.Muskan Mehra, Ishita Verma, Nihal Singh " Advancing Crime Analysis and Prediction: A Comprehensive Exploration of Machine Learning Applications in Criminal Justice" 2024.

8.Sameya Khatun, Kavyasree Banoth, Akshara Dilli, Soundarya Kakarlapudi, Sri Vani Karrola, G. Charles Babu "Machine Learning based Advanced Crime Prediction and Analysis" 2023.

9.Dr. P. Varaprasada Rao, Ganesh Koka, Sukesh Sunkari, "Prediction of Crime Data using Machine Learning Techniques" 2023.

10.R. Sathish Kumar, N.P. Saravanan, K. Nirmala Devi, P. Jayanthi. ""Empirical Analysis on Crime Prediction usingMachine Learning" 2023.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com