IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13597

AI-Driven Crime Prediction Using Machine Learning and Flask

Mrs.V. Anusha[1], Mrs.S. Sirisha[2]

Assistant professor, Department Of Computer And Science Of Engineering,

International school of technology and science for women, Rajamahendravaram, India[1]

M.Tech Student (246W1D5824), Department Of Computer And Science Of Engineering,

International school of technology and science for women, Rajamahendravaram, India[2]

Abstract: Crime analysis plays a vital role in ensuring public safety and effective law enforcement. The increasing rate of crimes makes it necessary to analyze large amounts of crime data efficiently. Traditional methods mainly depend on manual analysis of historical records, which are time-consuming and lack predictive capabilities. As a result, identifying crime patterns and trends becomes difficult. To overcome these limitations, a machine learning-based system is proposed. The system is developed using a Flask-based web application and processes crime datasets containing attributes such as crime type, location, and time. Data pre-processing techniques are applied to clean and prepare the data for analysis. Machine learning algorithms are then used to extract meaningful patterns from the data. Classification techniques help in predicting the type of crime, while clustering methods are used to identify crime-prone areas. The system also provides visualizations such as graphs and heatmaps, which make it easier to understand crime patterns and trends. By integrating data processing, prediction, and visualization into a single platform, the system improves decision-making and supports proactive crime prevention.

Keywords: Crime Analysis, Machine Learning, Crime Prediction, Flask, Data Pre-processing, Classification, Clustering, Data Visualization, HeatMaps

I.INTRODUCTION

1) Motivation

The motivation behind this project is the increasing number of crimes in society and the need for better public safety. Traditional crime analysis methods depend on manual work and historical records, which consume more time and effort. These methods are not efficient in predicting future crime trends. With the growth of digital data, there is a need for smart systems that can analyze large datasets quickly. Machine learning techniques help in identifying hidden patterns and relationships in crime data. By using these techniques, crime-prone areas and crime trends can be predicted effectively. This helps police departments in taking preventive actions and proper planning. Visualization tools such as graphs and heatmaps improve understanding of crime patterns. A web-based system also makes the application easy to access and use. The project aims to change crime analysis from reactive methods to proactive prevention methods. It supports better decision-making and efficient resource management. Ultimately, the system contributes to building safer and smarter communities.

2) Problem Statement

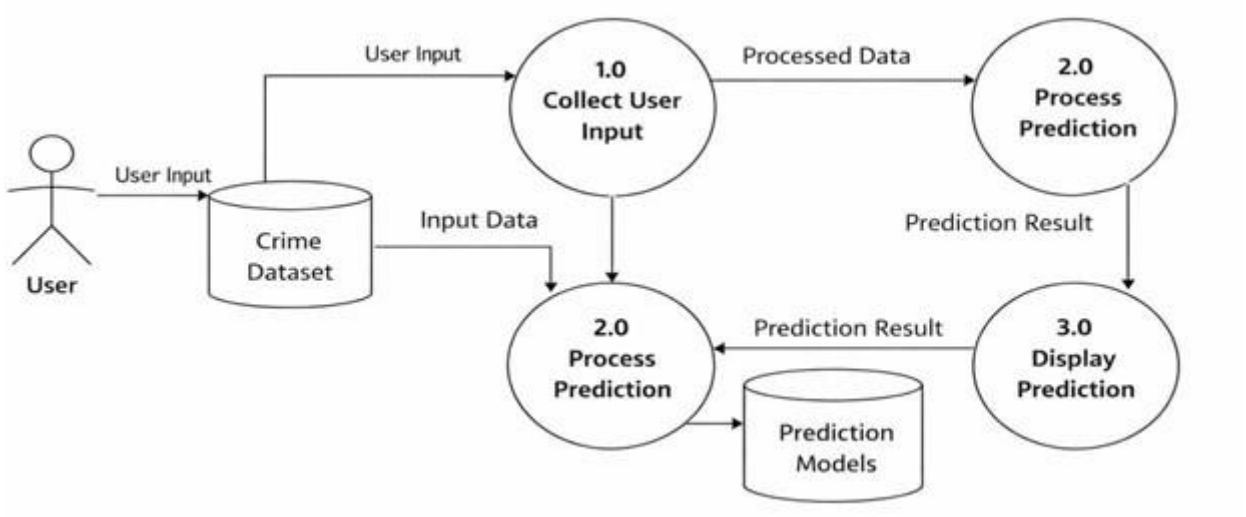
The increasing crime rate has become a major issue for society and law enforcement agencies. Large amounts of crime data are generated daily due to urbanization and population growth. Traditional methods of crime analysis mainly rely on manual processing of records. These methods are slow, time-consuming, and less efficient. They also fail to predict future crimes accurately. As a result, identifying hidden crime patterns becomes difficult. Detecting crime-prone areas and understanding criminal behaviour is limited in conventional systems. The absence of automated tools affects quick and proper decision-making. Authorities mostly take actions only after crimes occur. Therefore, there is a need for an intelligent and automated crime analysis system. Machine learning techniques can help in analyzing large datasets and predicting crime trends. Such a system can improve public safety and support proactive crime prevention.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13597



3) Objective of the Project

The main objective of this project is to develop a smart crime analysis and prediction system using machine learning techniques. The system aims to process large volumes of crime data efficiently. It helps in identifying patterns and trends from historical crime records. One objective is to predict the type of crime using classification algorithms. Another objective is to identify crime-prone areas using clustering methods. The project also aims to automate the crime analysis process and reduce manual effort. It improves the accuracy and speed of crime prediction. Visualization tools such as graphs and heatmaps are used for better understanding of crime data. The system is developed as a web-based application using Flask. It provides useful insights for law enforcement agencies to take preventive measures. The project supports better planning and decision-making. Overall, it aims to improve public safety and crime monitoring systems.

II. LITERATURE SURVEY

Crime prediction using machine learning has gained significant attention in recent years. A survey conducted in 2022 highlights various machine learning techniques such as Decision Tree, Support Vector Machine (SVM), and Neural Networks, comparing their performance based on dataset characteristics and accuracy requirements [1]. This study provides a foundational understanding of algorithm selection for crime prediction systems.

In 2023, another study explored machine learning approaches for crime prediction, offering a conceptual overview of the methods used [2]. However, it lacks implementation details relevant to web-based systems. Additionally, the dataset and methodology do not align with the requirements of the proposed model. The absence of visualization and real-time analysis further limits its applicability. Therefore, this study was not included in the final references.

A systematic review conducted in 2023 examined both machine learning and deep learning techniques for crime prediction [3]. While the study provides valuable comparisons, it mainly focuses on theoretical evaluation rather than practical system development. Furthermore, it does not support integration with web technologies, reducing its relevance to the proposed system. Hence, it was excluded from the final selection.

A separate study in 2024 presented a simple machine learning-based crime prediction model [4]. Although the approach is straightforward, it lacks innovation and does not support real-time data processing. The absence of scalability and flexibility makes it less suitable for practical applications. Additionally, visualization aspects are not adequately addressed, leading to its exclusion.

Another study published in 2024 proposed a system for crime analysis using machine learning techniques [5]. While it offers useful insights, it lacks advanced prediction mechanisms and does not support real-time processing. The visualization techniques are also limited, and the dataset used is not suitable for the proposed application. Therefore, this paper was excluded.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13597

A further 2024 study focused on crime prediction and trend analysis using machine learning [6]. However, it lacks integration with modern technologies and uses basic prediction models with lower accuracy. It also does not provide a detailed system architecture or support interactive visualization features. Due to these limitations, it was excluded in favour of more comprehensive studies.

Finally, a systematic literature review conducted in 2025 summarized recent advancements in crime prediction techniques [7]. However, it mainly compiles existing research without proposing a new model or providing technical implementation details. The lack of system architecture and integration aspects, particularly for web-based applications, reduces its relevance. Therefore, it was excluded from the final selection.

III.SYSTEM ANALYSIS

1 Existing System

The existing system for crime analysis mainly depends on traditional methods of collecting and examining historical crime data manually. Law enforcement agencies use basic statistical tools and record-based analysis to identify crime patterns and trends. These methods require significant human effort and consume a large amount of time, which reduces the overall efficiency of the system. Most existing systems focus only on analyzing past crime records and are unable to accurately predict future crime occurrences. They do not support real-time data processing and cannot efficiently manage large and continuously growing datasets. In addition, the absence of advanced machine learning algorithms limits the ability to identify hidden relationships and complex crime patterns within the data. Visualization support is also limited, making it difficult for users to understand crime trends clearly. Traditional systems are not integrated with modern web-based technologies, reducing accessibility and user interaction. Due to these limitations, existing systems are less effective in supporting proactive crime prevention and decision-making processes.

Disadvantages

The major disadvantages of the existing system are as follows:

a. The system depends heavily on manual data analysis, which is time-consuming and error-prone.
b. It does not support real-time crime monitoring and analysis.
c. Prediction accuracy is low due to the absence of advanced machine learning techniques.
d. The system mainly focuses on historical data and cannot forecast future crime trends effectively.
e. Visualization tools are limited and not interactive.
f. It cannot efficiently handle large-scale and continuously growing datasets.
g. The system lacks integration with modern web-based platforms.
h. Decision-making becomes reactive instead of proactive.

Proposed System

The proposed project is a Crime Analysis and Prediction System developed using machine learning techniques and a Flask-based web framework. The primary objective of this system is to analyze historical crime data and generate meaningful insights to predict future crime trends. The system uses advanced data processing methods to identify patterns based on factors such as crime type, location, and time. Initially, crime data is collected from reliable sources and then pre-processed by removing inconsistencies and handling missing values. After pre-processing, important features are selected to improve model performance. Machine learning algorithms such as classification and clustering techniques are then applied to train the model and identify hidden patterns in the dataset. Once trained, the system predicts possible future crime trends and provides accurate analytical insights. The proposed system also includes visualization features such as charts, graphs, and heatmaps, which help users understand crime distribution effectively. In addition, the project is integrated with a Flask-based web interface that improves accessibility and user interaction. By combining machine learning with interactive visualization, the proposed system provides an efficient and proactive solution for crime analysis and prevention.

Advantages of the Proposed System

The proposed system offers several advantages over traditional crime analysis methods. The system provides real-time crime analysis and prediction, enabling faster and better decision-making. Machine learning algorithms improve

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

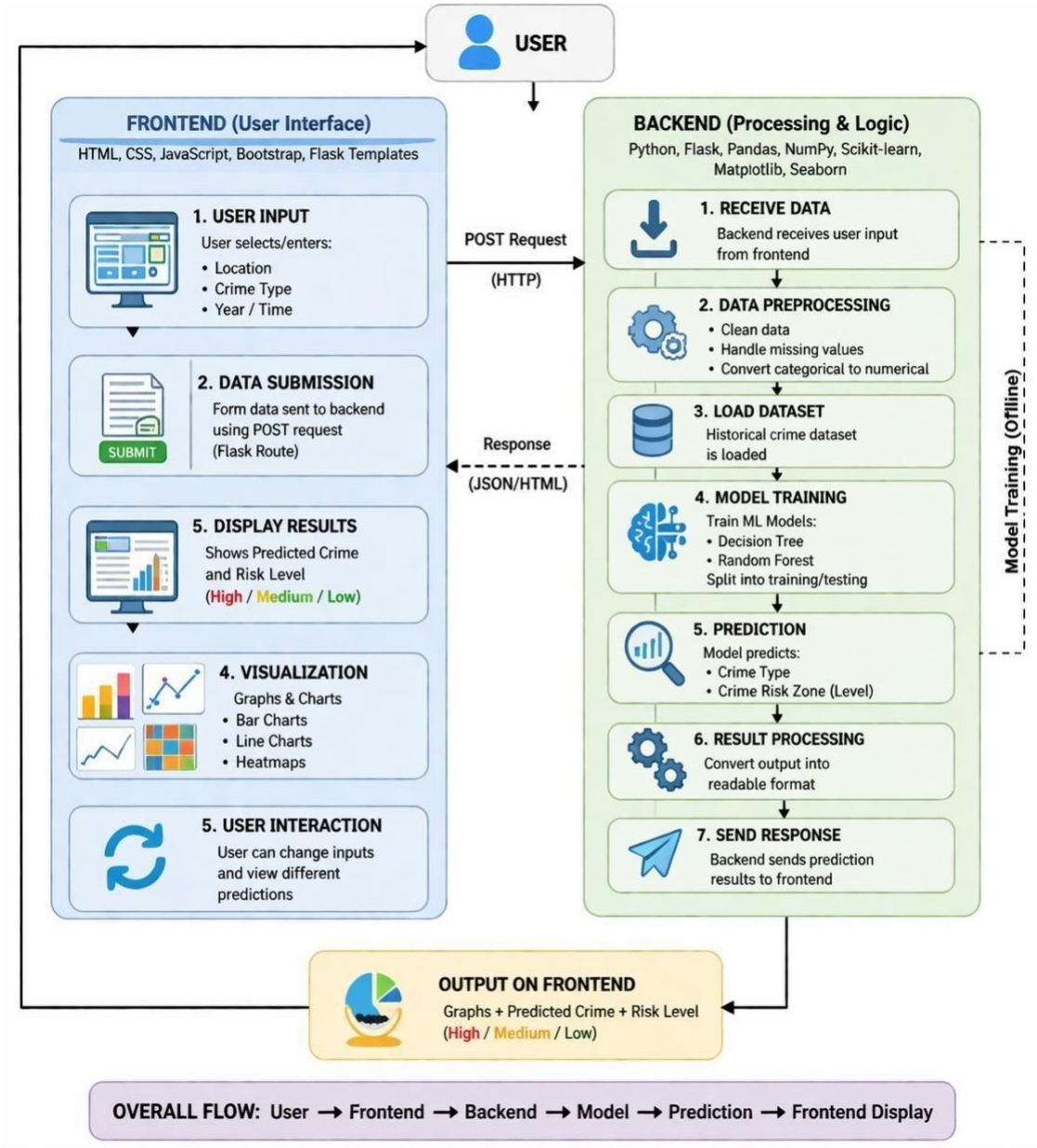
DOI: 10.17148/IARJSET.2026.13597

prediction accuracy by identifying hidden patterns in large datasets. The system can efficiently handle large volumes of data, making it scalable and reliable. It supports proactive crime prevention by forecasting future crime trends in advance. Interactive visualization tools such as charts, graphs, and heatmaps improve data understanding and interpretation. The Flask-based web interface makes the system user-friendly and easily accessible from different platforms. Automation reduces manual effort and minimizes human errors during analysis. The system also assists law enforcement agencies in better resource allocation and strategic planning. Furthermore, it integrates analysis, prediction, and visualization functionalities into a single platform and can be continuously updated with new data for improved performance.

Project Flow

The overall workflow of the system is illustrated in Figure 1, which represents the step-by-step process involved in crime analysis and prediction. The process begins with data collection, followed by data pre-processing and feature selection. After this, machine learning models are trained using the processed data. The trained model is then used for crime prediction and trend analysis. Finally, the results are visualized through charts, graphs, and heatmaps and displayed using a Flask-based web interface for better user interaction and accessibility.

Figure 1: Project Flowchart



IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

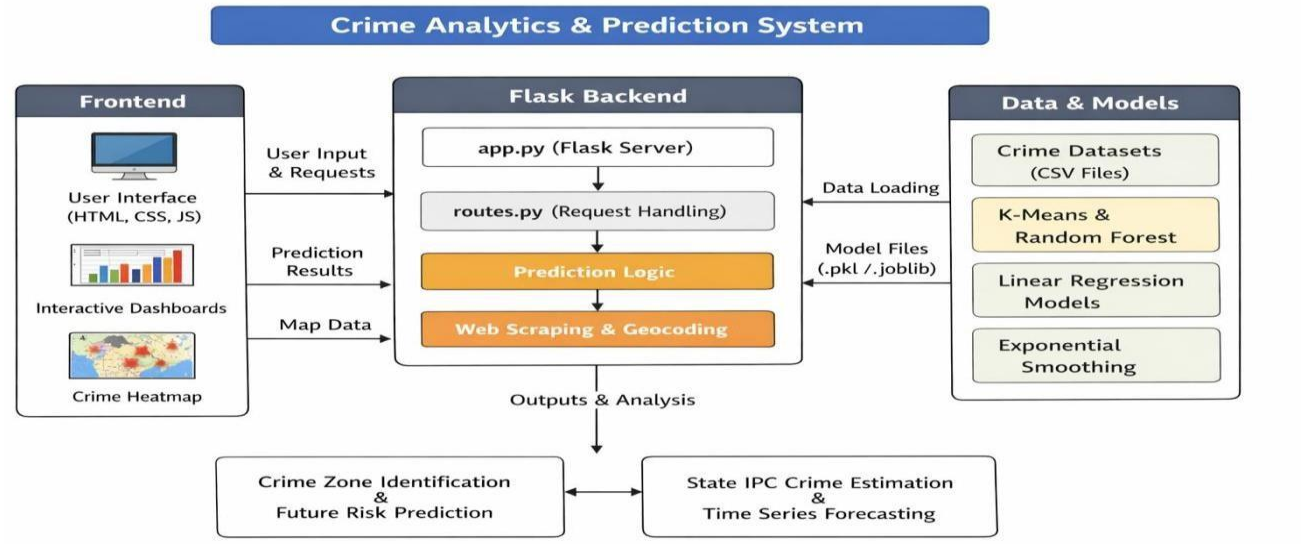
Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13597

IV.REQUIREMENT ANALYSIS

The Requirement Analysis phase of the Crime Analysis and Prediction System focuses on identifying the functional and non-functional requirements needed for efficient system performance. The functional requirements describe the major operations performed by the system. The system allows users to provide input details such as crime type, location, date, and time for prediction and analysis purposes. It also supports uploading historical crime datasets by users or administrators for further processing. Before analysis, the system performs preprocessing techniques such as handling missing values, removing duplicate records, and converting data into a suitable format. The system analyzes historical crime data to identify trends, patterns, and relationships between crime-related factors. Machine learning models such as Decision Tree and Random Forest are trained using processed datasets to improve prediction accuracy. Based on the trained models, the system predicts possible crime occurrences for specific locations and conditions. In addition, the system provides visualization features such as graphs, charts, and heatmaps to represent crime statistics clearly.

The non-functional requirements define the quality and reliability standards of the system. The system must provide high



performance by processing data quickly and generating predictions efficiently. Scalability is important so that the system can handle increasing amounts of crime data in the future. Reliability ensures that the system performs consistently and produces accurate results without failures. Usability focuses on providing a simple and user-friendly interface that can be easily operated by users and administrators. Security is also an essential requirement to protect crime-related information and prevent unauthorized access. The software requirements include Windows 10 operating system, Python programming language, Flask framework, HTML, CSS, JavaScript, and libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib. The hardware requirements include an Intel i3 processor, 8 GB RAM, 128 GB storage, keyboard, mouse, and monitor. The system architecture follows a layered design consisting of user interface, backend processing, and database/model layers, ensuring smooth communication between all system components.

V.SYSTEM DESIGN

The system design of the Crime Analysis and Prediction System focuses on creating an efficient, reliable, and user-friendly framework for handling crime data analysis and prediction. Input design plays a major role in collecting accurate and meaningful information from users. The system provides an interactive interface using dropdown menus, text fields, and selection controls that simplify data entry and reduce user errors. Validation techniques ensure that only valid states, districts, years, and crime-related information are accepted, improving the accuracy of machine learning predictions. The entered data is automatically converted into encoded values suitable for clustering, classification, and regression models. The system also supports real-time data collection from crime news sources to generate live heatmaps and trend analysis. Backend mechanisms handle missing or inconsistent inputs through fallback methods, ensuring smooth execution without

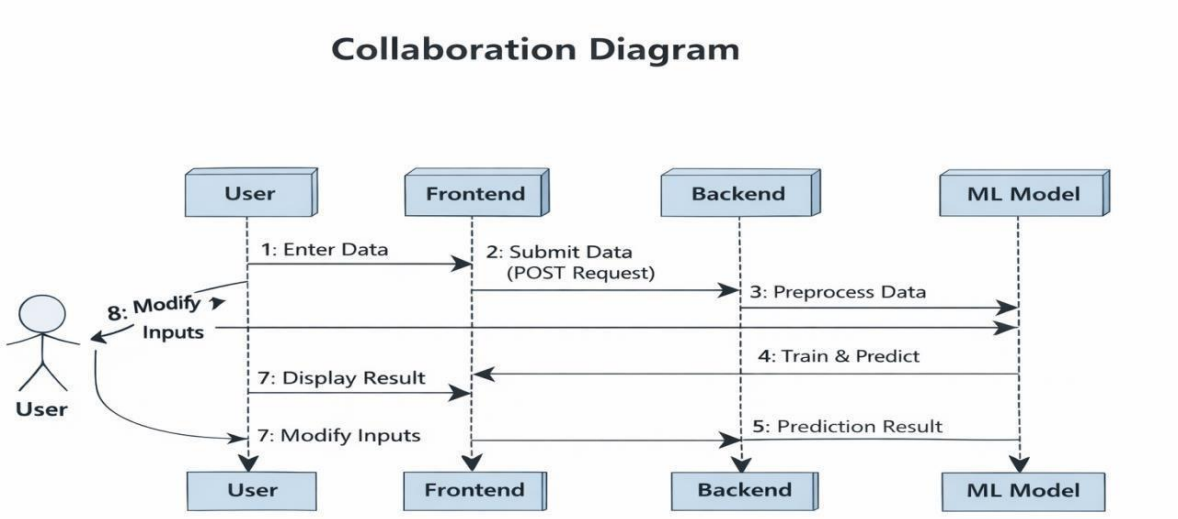
IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13597

system failure. Structured input processing improves efficiency and enables quick retrieval of historical datasets for analysis. The design is scalable, allowing the addition of new regions, years, and datasets without major modifications.



Output design focuses on presenting prediction results and crime analytics in a clear and understandable format. Charts, tables, dashboards, and color-coded maps are used to visualize crime intensity and trends effectively. Interactive dashboards allow users to filter information and analyze crime patterns across states and districts. Real-time crime headlines and updated heatmaps improve decision-making capabilities for law enforcement agencies and policymakers. The system maintains consistency in visualizations and labelling across all modules for better usability. UML diagrams such as Use Case, Class, Sequence, Collaboration, Deployment, Activity, Component, ER, and DFD diagrams are used to represent the structure, workflow, database relationships, and data movement within the system. These diagrams provide a complete blueprint for developing and understanding the software architecture of the Crime Analysis and Prediction System.

RESULTS

The proposed Crime Analysis and Prediction System was successfully implemented using various system modules, machine learning algorithms, and visualization techniques. The system was tested using historical crime datasets containing information such as crime type, location, date, and time of occurrence. The User Interface Module effectively handled user interactions by allowing users to enter details through input forms and displaying prediction results in a simple and understandable format. Charts, graphs, and heatmaps generated through the Visualization Module improved the interpretation of crime data and helped users identify crime-prone areas easily.

The Data Input and Validation Module ensured that all user inputs were accurate and properly formatted before processing. Invalid or incomplete entries were rejected using frontend and backend validation techniques. The Data Pre-processing Module cleaned the dataset by handling missing values, converting categorical values into numerical form, and normalizing the data for better model performance. The Dataset Management Module efficiently managed crime datasets stored in CSV and database formats, ensuring integrity and easy access during training and prediction.

The Machine Learning Module produced effective results in predicting crime risk levels and identifying crime patterns. Random Forest and Support Vector Machine (SVM) algorithms achieved high classification accuracy in predicting crime categories and risk levels. The Random Forest prediction model is represented by:

$$y = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \quad -$$

where ($h_t(x)$) represents the prediction of each decision tree and (T) represents the total number of trees. This ensemble approach improved prediction accuracy and reduced overfitting.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13597

The Support Vector Machine algorithm classified crime data by finding the optimal hyperplane using the equation

$$f(x) = w^T x + b$$

subject to:

$$y_i(w^T x_i + b) \geq 1$$

The clustering algorithms such as K-Means and DBSCAN effectively grouped similar crime locations and identified crime hotspots. K-Means clustering minimized the distance between data points and cluster centroids using the objective function

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

The DBSCAN algorithm identified dense crime regions based on neigh borhood density conditions

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

These clustering methods helped classify regions into high, medium, and low crime zones, enabling better crime monitoring and preventive planning.The Regression and Time-Series Forecasting techniques also produced valuable results. Linear Regression successfully estimated crime rates based on influencing factors such as population, location, and time using.ARIMA and LSTM models effectively analyzed crime trends over time and forecasted future crime occurrences. The LSTM model captured long-term dependencies in sequential crime data through forget gates, input gates, and output gates, improving future crime trend prediction accuracy.Geospatial heatmap analysis successfully visualized crime density across regions using Kernel Density Estimation (KDE):

$$f(x,y) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}, \frac{y - y_i}{h}\right)$$

The generated heatmaps clearly highlighted high-crime regions using colour variations, where red and orange represented high-risk areas and blue or green represented safer zones. These visualizations improved understanding of crime distribution and supported better decision-making for law enforcement authorities.The Result Processing Module converted prediction outputs into readable formats such as High, Medium, and Low risk levels. The Reporting Module successfully generated downloadable PDF and CSV reports containing prediction summaries and user activities. The System Management and Security Module ensured secure operation through user authentication, password hashing, error management, and activity logging.The developed system demonstrated good performance, high usability, and reliable prediction capability. The integration of machine learning algorithms, geospatial visualization, and forecasting techniques significantly enhanced crime analysis efficiency. Overall, the project successfully achieved its objective of predicting crime patterns, identifying hotspots, and assisting authorities in taking preventive actions through accurate analysis and visualization.

CONCLUSION

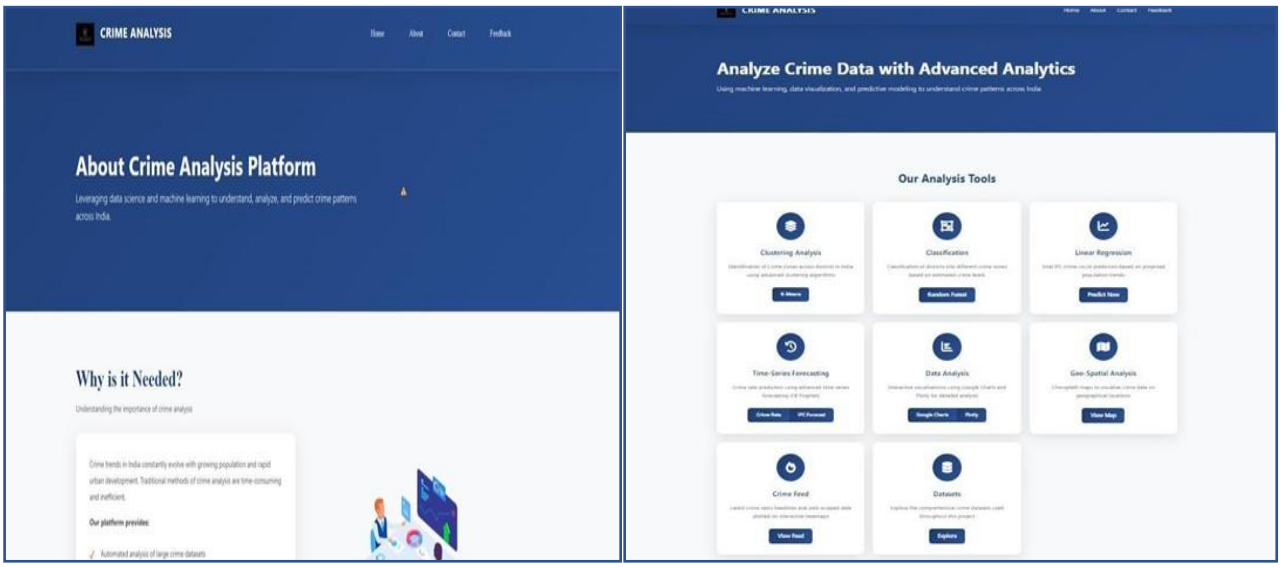
The Crime Analysis and Prediction System is a web-based application developed to analyze historical crime data, predict crime types, and identify risk levels for different locations and time periods. The system successfully integrates a user-friendly frontend with a powerful backend that performs data pre-processing, machine learning-based prediction, and data visualization.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.13597



Algorithms such as Decision Tree, Random Forest, Linear Regression, K-Means Clustering, and Exponential Smoothing help generate accurate and meaningful predictions. Extensive testing, including unit, integration, functional, and acceptance testing, ensured the reliability and efficiency of the system. Visual tools such as charts, heatmaps, and reports make crime information easy to understand and interpret. The project effectively supports crime awareness, monitoring, and decision-making processes. Future enhancements can include real-time crime data integration, advanced AI and deep learning models, mobile application support, GIS-based mapping, personalized dashboards, predictive policing features, multilingual support, and cloud-based big data analytics. These improvements will further increase the accuracy, scalability, and practical usefulness of the system in enhancing public safety and crime prevention.

REFERENCES

[1]. A. K. Singh and P. K. Gupta, "Crime Prediction Using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 179, no. 21, pp. 1–5, 2018.

[2]. S. Nath, "Crime Pattern Detection Using Data Mining," in *Proceedings of the IEEE International Conference on Electronics, Communication and Aerospace Technology*, 2017.

[3]. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[4]. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[5]. M. Grinberg, *Flask Web Development*, 2nd ed., O'Reilly Media, 2018.

[6]. J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2012.

[7]. P. V. Kumar et al., "Crime Data Analysis Using Clustering Techniques," *International Journal of Advanced Research in Computer Science*, 2019.

[8]. D. Pedregosa et al., "Machine Learning Approaches for Crime Data Analysis," *IEEE Access*.