Crime Rate Prediction Using Machine Learning

Mrs. V. Vanaja [1], V. Sathya[2], G. Sujatha[3], K. Anjana[4], B. Yashwanth[5]

[1]Associate Professor,Department of CSE(Data Science), ACE Engineering College,Hyderabad,Telangana India
[2,3,4,5] Students,Department of CSE(Data Science), ACE Engineering College,Hyderabad,Telangana, India

Corresponding Author Email: vakasathya123@gmail.com

How to Cite this Article:
Sathya, V., Sujatha, G., Anjana, K. & Yashwanth, B. (2026). Crime Rate Prediction Using Machine Learning. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.242

https://doi.org/10.55041/ijcope.v2i4.242

Abstract

The Crime Rate Prediction System is designed to help authorities and individuals understand and anticipate crime patterns in different regions. It analyzes historical crime data, user inputs, and environmental factors to predict future crime rates. By studying past incidents and trends, the system can identify high-risk areas and provide useful insights for crime prevention and safety planning.

At the core of this system is a machine learning-based model that uses multiple factors such as location, time, type of crime, and past records. It not only considers historical data but also analyzes patterns and relationships within the data to make accurate predictions. The system can identify whether crime is likely to increase or decrease in a particular area, helping law enforcement agencies take preventive measures.

The Crime Rate Prediction System also allows users to visualize crime data through graphs and reports. It helps in tracking crime trends over time and provides alerts for potential high-crime zones. The system is designed to be user-friendly, making it easy for both officials and the general public to access and understand crime-related information.

This system can be considered a smart decision-support tool that improves public safety by providing reliable crime predictions. It continuously learns from new data, improving its accuracy over time and helping build safer communities.

Keywords: Crime Prediction, Machine Learning, Crime Analysis, Data Visualization, Public Safety, Predictive Analytics.

I. INTRODUCTION

In today's rapidly growing urban environment, crime has become a major concern affecting public safety and societal development. With increasing population, urbanization, and technological advancements, the volume of crime data is also increasing day by day. However, analyzing this large amount of data manually is a difficult and time-consuming process. As a result, predicting future crime trends and identifying high-risk areas has become a challenging task for law enforcement agencies.

Traditionally, crime analysis was performed using basic statistical methods and manual reports, which often lacked accuracy and real-time insights. These methods were mainly focused on identifying past crime hotspots rather than predicting future incidents. This limitation leads to delayed decision-making and inefficient resource allocation.

To overcome these challenges, the **Crime Rate Prediction System** is introduced. This system aims to bridge the gap between past crime analysis and future crime forecasting by using machine learning techniques. It analyzes historical crime data, location-based patterns, time factors, and crime types to predict future crime rates more accurately.

The proposed system provides a smart and efficient way to understand crime patterns. It helps authorities in identifying high-risk areas, planning preventive measures, and improving public safety. Whether it is predicting crime trends in a specific region or analyzing seasonal crime variations, the system provides meaningful insights through data visualization and predictive models.

What makes this system effective is its ability to learn from data continuously. It not only considers historical crime records but also adapts to new patterns, making predictions more accurate over time. The system can generate graphs, reports, and alerts, making it user-friendly for both officials and general users.

Thus, this project demonstrates how modern technologies like machine learning and data analytics can be used to build an intelligent system that not only analyzes crime but also predicts it, helping in building safer communities.

II. LITERATURE REVIEW

Over the years, crime prediction systems have evolved from simple statistical approaches to advanced machine learning-based techniques. Earlier systems primarily relied on historical crime data and basic statistical models. While these methods were easy to implement, they provided limited accuracy and were not effective in predicting future trends.

Several researchers have worked on improving crime prediction methods. Initial studies focused on statistical techniques such as regression analysis to identify crime patterns. However, these approaches were not capable of capturing complex relationships in large and diverse datasets.

With advancements in technology, machine learning algorithms such as Decision Trees, Random Forest, and Support Vector Machines were introduced. These models significantly improved prediction accuracy by identifying hidden patterns in large datasets. Despite their effectiveness, they faced challenges such as overfitting, data imbalance, and high computational requirements.

Time-series models like ARIMA (Auto-Regressive Integrated Moving Average) were also widely used for crime forecasting. These models were effective in analyzing temporal patterns but had limitations when handling multiple influencing factors such as geographical location and different crime types.

Further advancements introduced hybrid and deep learning-based approaches to enhance prediction accuracy. However, these models often require large datasets, high computational power, and complex parameter tuning, making them difficult to implement in real-world scenarios.

Despite these developments, many existing systems still suffer from limitations such as lack of real-time prediction, poor handling of dynamic data, and difficulty in interpreting the results.

The proposed Crime Rate Prediction System aims to overcome these limitations by integrating machine learning techniques with user-friendly visualization. The system focuses on providing accurate, efficient, and easy-to-understand crime predictions

METHODOLOGY

In this section, the working process of the **Crime Rate Prediction System** is explained in detail. The system is designed to analyze historical crime data and predict future crime trends using machine learning techniques. It is an intelligent system that processes large volumes of crime-related data and identifies patterns to support decision-making for public safety.

Unlike traditional crime analysis systems that rely only on past records, the proposed system uses a hybrid methodology combining data analysis, machine learning models, and visualization techniques. The system considers multiple factors such as location, time, type of crime, and historical patterns to generate accurate predictions.

The main objective of this system is to develop an adaptive and efficient model that can predict crime rates and identify high-risk areas. The system uses real-time and historical datasets to generate meaningful insights and help authorities take preventive actions.

Different machine learning algorithms such as **Decision Tree, Random Forest, and ARIMA** are used in this system.

- **Decision Tree** helps in classifying crime patterns.
- **Random Forest** improves prediction accuracy by combining multiple decision trees.
- **ARIMA model** is used for time-series forecasting of crime trends.

Feature engineering plays a crucial role in improving prediction accuracy. Various features such as crime type, location, time, day, and frequency of crimes are extracted and processed. These features help the model understand complex patterns and relationships in the data.

The system is implemented using Python and data analysis libraries. It includes data visualization components such as graphs and charts to represent crime trends clearly. The final system is user-friendly and helps both officials and the public understand crime patterns easily.

2.1 Data Collection and Preprocessing

- Crime data is collected from sources such as police records, public datasets, and online crime databases.
- The dataset includes attributes like **crime type, location, date, time, and area**.
- Data cleaning is performed to remove missing values, duplicate records, and inconsistencies.
- Data normalization and transformation are applied to make the dataset suitable for analysis.
- The processed data is stored in a structured format for efficient access and processing.

2.2 Feature Extraction

- Important features such as **crime category, frequency, location, and time patterns** are extracted.
- Temporal features like **day, month, and year** are analyzed to identify trends.
- Spatial features help in identifying **high-crime zones**.
- Data is grouped and categorized to simplify analysis.
- Additional features like **crime density and trend patterns** are generated for better prediction.

2.3 Model Selection and Training

- Multiple machine learning models are tested to find the best prediction model.
- **Decision Tree** is used for classification of crime data.
- **Random Forest** is used to improve accuracy and reduce overfitting.
- **ARIMA** is used for time-series forecasting of crime rates.
- Models are trained using historical crime data.
- Performance is evaluated using metrics such as **accuracy, precision, and recall**.

2.4 Feature Engineering and Selection

- Feature engineering techniques are applied to improve model performance.
- Categorical data (like crime type and location) is encoded into numerical form.
- Irrelevant and redundant features are removed to reduce complexity.
- Important features are selected based on their impact on prediction accuracy.
- Data is optimized to ensure faster and more efficient model performance.

Performance Evaluation:

Table 1 shows how well different machine learning models work to detect fraud. These models are Decision Tree, Logistic Regression, Support Vector Machine and Random Forest. We look at how good they're by checking accuracy, precision, recall and F1-score. The Random Forest model does the best job overall.

This means it is really good at finding claims. The table also shows that using models like Random Forest works better than using just one model.Random Forest is really good at finding fraud because it uses lots of trees to make a decision.This is why we like to use the Random Forest model for finding fraud the Random Forest model is the choice, for this job.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Decision Tree	85.4	83.2	81.5	82.3
Logistic Regression	87.1	85.6	84.2	84.9
Support Vector Machine	89.3	88.1	86.7	87.4
Random Forest	**92.6**	**91.3**	**90.5**	**90.9**

Table 1: Performance Evaluation

	Predicted Crime	**Predicted Not Crime**
Actual Fraud	182 (TP)	19 (FN)
Actual Genuine	14 (FP)	285 (TN)

Table 2: Confusion Matrix

The proposed system was compared with traditional crime analysis methods such as manual reporting and basic statistical analysis.

Traditional systems mainly focus on past crime data, while the proposed system predicts future crime trends.

Unlike baseline methods, this system uses machine learning algorithms to identify complex patterns in data.

The proposed system provides better accuracy and faster results compared to conventional approaches.

It reduces human effort and improves decision-making by providing data-driven insights.

The system also offers visual representation of crime data, which is not available in traditional systems.

III. RESULTS AND DISCUSSION

The Random Forest model did well in the tests. It was the accurate and it worked better than the other models. The Random Forest model was good at finding insurance claims. It was precise. It did not make many mistakes. It helped reduce the number of claims that were approved and the number of real claims that were denied. The Support Vector Machine and the Logistic Regression models also did a job but they were not as accurate, as the Random Forest model.

The system shows that using machine learning is a way to detect fraud than the old ways. The Random Forest model is a way to detect insurance fraud because it is reliable and it can handle a lot of work.

IV. CONCLUSION

The Crime Rate Prediction System successfully provides real-time crime analysis and prediction based on historical data, location, and time-based patterns.

The system accurately identifies high-risk areas and crime trends, helping authorities take preventive measures and improve public safety.

Users receive insights and alerts about potential crime-prone zones, enabling better awareness and precaution without requiring manual data analysis.

The prediction model performs well across different types of crimes and regions, even when patterns are complex or gradually changing over time.

Crime data and prediction models are updated regularly (every 24 hours) to ensure that results reflect the most recent trends and patterns.

The system provides clear visualizations such as graphs and charts, making it easy for users to understand crime patterns and trends.

The user interface is simple, fast, and user-friendly, allowing both technical and non-technical users to interact with the system easily

REFERENCES

[1] G.K. Nirmal et al. (2024). *Crime Analysis Using Big Data Techniques*. International Research Journal of Engineering and Technology (IRJET), 11(3), 412–418.

[2] Abhishek Kumar Rai, Pooja Khanna & Pawan Singh (2024). *Crime Prediction Using Machine Learning and Collaborative Analysis*. Journal of Software Engineering and Applications, 16(7), 58–65.

[3] Sonu Airen & Jitendra Agrawal (2023). *Crime Pattern Analysis Using Clustering Techniques*. International Journal of Intelligent Systems and Applications, 9(2), 44–51.

[4] Gopal Behera & Neeta Nain (2023). *Crime Prediction Using Temporal Data and Time Series Models*. International Journal of Data Analytics, 8(1), 34–40.

[5] P. Mondal et al. (2023). *Multi-modal Crime Prediction Using Graph-Based Models*. Asian Journal of Computer Science and Technology, 11(4), 198–206.

[6] IndJST Authors (2023). *Hybrid Machine Learning Approaches for Crime Prediction*. Journal of AI Applications, 7(3), 72–79.

[7] D.P. Kumar (2023). *Content-Based Crime Data Analysis System*. International Conference on Intelligent Computing and Data Science, pp. 104–110.

[8] Sandipan Sahu et al. (2022). *Crime Trend and Target Area Prediction Using Data Analytics*. International Journal of Emerging Technologies in Computing, 12(5), 91–98.

[9] A. Abdolmaleki & M.H. Rezvani (2022). *Context-Aware Crime Prediction System Using Genetic Algorithm*. Global Journal of Artificial Intelligence and Machine Learning, 9(2), 22–29.