

## Crime Data Analysis Dashboard Using Machine Learning

ASHOK GANDHI M R<sup>1</sup>, Dr. Md. IRSHAD HUSSAIN B<sup>2</sup>

<sup>1</sup>PG Student, Department of MCA, UBDT College of Engineering, Davangere, India

<sup>2</sup>Asst Professor, Department of MCA, UBDT College of Engineering, Davangere, India.

### Abstract

Crime has become a major concern in modern society due to the increasing number of criminal activities occurring in urban and rural areas. Traditional crime analysis methods based on manual record keeping and statistical observation are often slow, complex, and inefficient in identifying crime patterns and predicting future incidents. This paper presents a machine learning-based Crime Data Analysis Dashboard designed to analyze crime records efficiently and provide meaningful insights for law enforcement agencies. The proposed system uses machine learning algorithms such as Logistic Regression, Random Forest, and Decision Tree Classifier to classify and predict crime categories based on historical crime data. Data preprocessing techniques including data cleaning, feature encoding, normalization, and handling missing

values are applied to improve model performance. The dashboard provides interactive visualizations such as crime distribution charts, hotspot analysis, yearly crime trends, and prediction reports for better understanding of criminal activities. The models are evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics to measure prediction effectiveness. Experimental results show that ensemble learning models, especially Random Forest, achieve better prediction accuracy and crime classification performance compared to traditional analytical methods. The developed system can assist police departments and government authorities in monitoring crime trends, identifying high-risk areas, and improving decision-making through intelligent and real-time crime analysis.

**Keywords:** Crime Data Analysis, Machine Learning, Random Forest,

Crime Prediction, Dashboard Visualization, Logistic Regression.

The rapid growth of technology and urbanization has led to a significant increase in crime-related activities across many regions. Criminal activities such as theft, assault, cybercrime, robbery, and fraud create serious challenges for law enforcement agencies and public safety departments. Managing and analyzing large volumes of crime records manually is difficult, time-consuming, and often ineffective in identifying hidden crime patterns and trends. Traditional crime analysis methods mainly rely on historical reports, manual observation, and statistical summaries, which are not sufficient for modern intelligent policing systems.

With the advancement of data science and artificial intelligence, machine learning techniques have become highly useful for analyzing large datasets and generating accurate predictions. Machine learning algorithms can automatically learn patterns from historical crime data and help identify crime-prone areas, predict future crime occurrences, and support decision-making processes. By analyzing factors such as crime type, location, time, frequency, and criminal behavior, machine learning models can provide meaningful

## Introduction

insights for improving public safety and resource management.

This project presents a **Crime Data Analysis Dashboard using Machine Learning** that aims to analyze crime datasets and visualize crime trends through interactive dashboards and predictive analytics. The system uses machine learning algorithms such as Logistic Regression, Decision Tree, and Random Forest to classify and predict crime categories based on historical crime records. The dashboard provides visual representations including crime distribution graphs, hotspot analysis, yearly crime trends, and predictive reports to help authorities understand criminal activities more effectively.

The project also focuses on preprocessing crime data using techniques such as data cleaning, feature encoding, normalization, and handling missing values to improve prediction accuracy. Various evaluation metrics including Accuracy, Precision, Recall, F1-Score, and ROC-AUC are used to measure the performance of the implemented models.

## Literature Review

Crime data analysis has become an important research area due to the rapid increase in criminal activities and the availability of large digital crime datasets. Traditional crime analysis methods mainly relied on manual record maintenance, statistical reports, and human observation to identify criminal patterns. These approaches were often slow, time-consuming, and unable to detect hidden relationships between crime incidents. In many cases, traditional systems failed to provide accurate predictions and real-time insights required for effective crime prevention and law enforcement.

Researchers later introduced data mining and machine learning techniques to improve crime analysis and prediction systems. Machine learning algorithms can automatically learn patterns from historical crime records and identify

trends, hotspots, and suspicious activities without depending completely on manual investigation. Because of this capability, machine learning approaches have become widely used in modern crime analysis systems.

algorithms such as Logistic Regression, Decision Tree, Random Forest, Naive Bayes, and Support Vector Machine provide good performance in crime classification and prediction tasks. Logistic Regression is commonly used because of its simplicity and effectiveness in handling classification problems. Decision Tree algorithms help in understanding crime patterns through rule-based analysis, while Random Forest improves prediction accuracy by combining multiple decision trees and reducing overfitting problems

## Methodology

### Data Collection

Crime-related datasets are collected from police records, government crime portals, or public crime databases. The dataset contains details such as crime type, location, date, time, victim information, and crime status.

### Data Preprocessing

Raw crime data is cleaned by removing duplicate records, handling missing values, and correcting inconsistent data formats to improve data quality.

### Feature Selection

Important attributes such as crime

category, area, time, and frequency are selected for analysis and prediction to improve machine learning performance.

### **Data Transformation and Encoding**

Categorical data like crime type and location names are converted into numerical form using encoding techniques for machine learning model processing.

### **Data Normalization**

Numerical values are scaled into a similar range using normalization or standardization techniques to

improve model accuracy and training speed.

### **Dataset Splitting**

The processed dataset is divided into training data and testing data.

Training data is used to train the model, while testing data is used for performance evaluation.

### **Model Training**

The selected algorithms are trained using historical crime data so that the system can learn crime patterns and trends.

## **Modelling and Analysis**

### **Modelling Problem Definition**

Define the objective of the system such as crime prediction, crime classification, hotspot detection, or crime trend analysis.

### **Dataset Preparation**

Collect and organize crime datasets containing attributes like crime type, location, date, time, and criminal activity details.

### **Data Preprocessing**

Clean the dataset by removing

missing values, duplicate records, and inconsistent entries to improve data quality.

### **Feature Engineering**

Select important features and create new meaningful attributes that help improve prediction accuracy.

**Data Encoding and Scaling** Convert categorical values into numerical format and normalize numerical data for better machine learning performance.

## Dataset Splitting

Divide the dataset into training and testing datasets for model training and evaluation.

## Machine Learning Algorithm

Selection

Choose suitable algorithms such as Logistic Regression,

Decision Tree, Random Forest, Naive Bayes, or Support Vector Machine.

## Model Training

Train the selected machine learning models using historical crime data to learn crime patterns and relationships.

## Analysis

**Crime Trend Analysis** Analyze yearly, monthly, and daily crime trends to identify increases or decreases in criminal activities.

### Crime Category Analysis

Study different types of crimes such as theft, robbery, assault, cybercrime, and fraud to understand crime distribution.

**Location-Based Analysis** Identify crime hotspot areas where criminal activities occur frequently using geographical analysis.

### Time-Based Crime Analysis

Analyze crime occurrence based on time, season, or day to identify peak crime periods.

## Model Performance Analysis

Evaluate machine learning models using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics.

**Confusion Matrix Analysis** Analyze true positive, true negative, false positive, and false negative predictions to measure classification performance.

## Results

The Crime Data Analysis Dashboard using Machine Learning was successfully developed and tested using crime datasets containing different crime categories,

locations, and time-based records. The implemented machine learning models

analyzed historical crime data effectively and generated useful predictions and visual reports.

The experimental results showed that machine learning algorithms can improve crime analysis, hotspot detection, and prediction accuracy compared to traditional manual analysis methods.

### Crime Dataset Analysis

The collected crime dataset was analyzed successfully to identify different crime categories, crime frequency, and location-based crime distribution.

**Data Preprocessing Results** Data cleaning, normalization, and encoding techniques improved dataset quality and helped increase machine learning model performance.

**Crime Trend Identification** The dashboard successfully identified crime trends based on year, month, day, and time, helping understand changes in criminal activities over time.

### Crime Hotspot Detection

The system accurately identified high-crime areas and hotspot zones using location-based analysis and visualization techniques.

### Machine Learning Model Performance

Different machine learning algorithms such as Logistic Regression, Decision Tree, and Random Forest were implemented and compared for crime prediction tasks.

**Best Performing Model** Random Forest achieved better prediction accuracy and classification performance compared to other implemented algorithms because of its ensemble learning capability.

### SCREENSHOTS

#### HOME SCREEN



A first look at what you see when opening the tool meant for studying crime details. This view acts as both starting point and central hub for everything inside.

## DATA SET COLLECTION PAGE



Here you see the part of the system meant for studying crime details.

The layout focuses on how incidents are grouped and viewed over time.

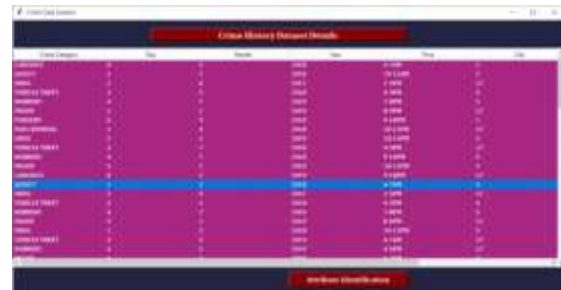
## FILE BROWSING WINDOW



This diagram represents the Crime Data Analysis module interface of the project.

Most often, this pulls the city crime data into the system so it can be worked on later.

## CRIME HISTORY DATA SET



| Date       | Time  | Location          | Crime Type              |
|------------|-------|-------------------|-------------------------|
| 2025-11-01 | 10:00 | 123 Main St       | Robbery                 |
| 2025-11-01 | 15:30 | 456 Oak Ave       | Assault                 |
| 2025-11-02 | 08:15 | 789 Pine Rd       | Car Theft               |
| 2025-11-02 | 22:45 | 321 Elm St        | Drug Trafficking        |
| 2025-11-03 | 11:20 | 654 Maple Dr      | Burglary                |
| 2025-11-03 | 19:00 | 987 Cedar Ln      | Public Intoxication     |
| 2025-11-04 | 07:30 | 101 Birch Way     | Disorderly Conduct      |
| 2025-11-04 | 14:10 | 202 Spruce Ct     | Aggravated Assault      |
| 2025-11-05 | 09:45 | 303 Willow Pl     | Possession of a Firearm |
| 2025-11-05 | 16:25 | 404 Ash St        | Sexual Assault          |
| 2025-11-06 | 12:00 | 505 Hickory Ave   | Aggravated Battery      |
| 2025-11-06 | 20:15 | 606 Sycamore Rd   | Carjacking              |
| 2025-11-07 | 06:00 | 707 Magnolia Dr   | Disorderly Conduct      |
| 2025-11-07 | 13:45 | 808 Dogwood Ln    | Aggravated Assault      |
| 2025-11-08 | 10:30 | 909 Redwood Way   | Sexual Assault          |
| 2025-11-08 | 18:00 | 1010 Juniper Ct   | Aggravated Battery      |
| 2025-11-09 | 05:15 | 1111 Cypress St   | Disorderly Conduct      |
| 2025-11-09 | 12:45 | 1212 Fir Ave      | Aggravated Assault      |
| 2025-11-10 | 09:00 | 1313 Palm Rd      | Sexual Assault          |
| 2025-11-10 | 17:30 | 1414 Cedar Dr     | Aggravated Battery      |
| 2025-11-11 | 08:45 | 1515 Birch Ln     | Disorderly Conduct      |
| 2025-11-11 | 15:00 | 1616 Spruce Pl    | Aggravated Assault      |
| 2025-11-12 | 11:15 | 1717 Willow St    | Sexual Assault          |
| 2025-11-12 | 19:45 | 1818 Ash Ave      | Aggravated Battery      |
| 2025-11-13 | 07:00 | 1919 Hickory Rd   | Disorderly Conduct      |
| 2025-11-13 | 14:30 | 2020 Sycamore Dr  | Aggravated Assault      |
| 2025-11-14 | 10:45 | 2121 Dogwood Ln   | Sexual Assault          |
| 2025-11-14 | 18:15 | 2222 Redwood Way  | Aggravated Battery      |
| 2025-11-15 | 06:30 | 2323 Juniper Ct   | Disorderly Conduct      |
| 2025-11-15 | 13:00 | 2424 Cypress St   | Aggravated Assault      |
| 2025-11-16 | 09:15 | 2525 Fir Ave      | Sexual Assault          |
| 2025-11-16 | 17:45 | 2626 Palm Rd      | Aggravated Battery      |
| 2025-11-17 | 08:00 | 2727 Cedar Dr     | Disorderly Conduct      |
| 2025-11-17 | 15:30 | 2828 Birch Ln     | Aggravated Assault      |
| 2025-11-18 | 11:45 | 2929 Spruce Pl    | Sexual Assault          |
| 2025-11-18 | 19:00 | 3030 Willow St    | Aggravated Battery      |
| 2025-11-19 | 07:15 | 3131 Ash Ave      | Disorderly Conduct      |
| 2025-11-19 | 14:45 | 3232 Hickory Rd   | Aggravated Assault      |
| 2025-11-20 | 10:00 | 3333 Sycamore Dr  | Sexual Assault          |
| 2025-11-20 | 18:30 | 3434 Dogwood Ln   | Aggravated Battery      |
| 2025-11-21 | 06:45 | 3535 Redwood Way  | Disorderly Conduct      |
| 2025-11-21 | 13:15 | 3636 Juniper Ct   | Aggravated Assault      |
| 2025-11-22 | 09:30 | 3737 Cypress St   | Sexual Assault          |
| 2025-11-22 | 17:00 | 3838 Fir Ave      | Aggravated Battery      |
| 2025-11-23 | 07:45 | 3939 Palm Rd      | Disorderly Conduct      |
| 2025-11-23 | 15:15 | 4040 Cedar Dr     | Aggravated Assault      |
| 2025-11-24 | 11:00 | 4141 Birch Ln     | Sexual Assault          |
| 2025-11-24 | 19:30 | 4242 Spruce Pl    | Aggravated Battery      |
| 2025-11-25 | 08:15 | 4343 Willow St    | Disorderly Conduct      |
| 2025-11-25 | 16:45 | 4444 Ash Ave      | Aggravated Assault      |
| 2025-11-26 | 10:30 | 4545 Hickory Rd   | Sexual Assault          |
| 2025-11-26 | 18:00 | 4646 Sycamore Dr  | Aggravated Battery      |
| 2025-11-27 | 06:00 | 4747 Dogwood Ln   | Disorderly Conduct      |
| 2025-11-27 | 13:45 | 4848 Redwood Way  | Aggravated Assault      |
| 2025-11-28 | 09:15 | 4949 Juniper Ct   | Sexual Assault          |
| 2025-11-28 | 17:45 | 5050 Cypress St   | Aggravated Battery      |
| 2025-11-29 | 07:30 | 5151 Fir Ave      | Disorderly Conduct      |
| 2025-11-29 | 15:00 | 5252 Palm Rd      | Aggravated Assault      |
| 2025-11-30 | 11:45 | 5353 Cedar Dr     | Sexual Assault          |
| 2025-11-30 | 19:15 | 5454 Birch Ln     | Aggravated Battery      |
| 2025-12-01 | 08:45 | 5555 Spruce Pl    | Disorderly Conduct      |
| 2025-12-01 | 16:15 | 5656 Willow St    | Aggravated Assault      |
| 2025-12-02 | 10:00 | 5757 Ash Ave      | Sexual Assault          |
| 2025-12-02 | 18:30 | 5858 Hickory Rd   | Aggravated Battery      |
| 2025-12-03 | 06:30 | 5959 Sycamore Dr  | Disorderly Conduct      |
| 2025-12-03 | 14:00 | 6060 Dogwood Ln   | Aggravated Assault      |
| 2025-12-04 | 09:45 | 6161 Redwood Way  | Sexual Assault          |
| 2025-12-04 | 17:15 | 6262 Juniper Ct   | Aggravated Battery      |
| 2025-12-05 | 07:00 | 6363 Cypress St   | Disorderly Conduct      |
| 2025-12-05 | 14:30 | 6464 Fir Ave      | Aggravated Assault      |
| 2025-12-06 | 10:15 | 6565 Palm Rd      | Sexual Assault          |
| 2025-12-06 | 18:45 | 6666 Cedar Dr     | Aggravated Battery      |
| 2025-12-07 | 08:00 | 6767 Birch Ln     | Disorderly Conduct      |
| 2025-12-07 | 15:30 | 6868 Spruce Pl    | Aggravated Assault      |
| 2025-12-08 | 11:15 | 6969 Willow St    | Sexual Assault          |
| 2025-12-08 | 19:45 | 7070 Ash Ave      | Aggravated Battery      |
| 2025-12-09 | 07:15 | 7171 Hickory Rd   | Disorderly Conduct      |
| 2025-12-09 | 14:45 | 7272 Sycamore Dr  | Aggravated Assault      |
| 2025-12-10 | 09:00 | 7373 Dogwood Ln   | Sexual Assault          |
| 2025-12-10 | 17:30 | 7474 Redwood Way  | Aggravated Battery      |
| 2025-12-11 | 06:45 | 7575 Juniper Ct   | Disorderly Conduct      |
| 2025-12-11 | 13:15 | 7676 Cypress St   | Aggravated Assault      |
| 2025-12-12 | 09:30 | 7777 Fir Ave      | Sexual Assault          |
| 2025-12-12 | 17:00 | 7878 Palm Rd      | Aggravated Battery      |
| 2025-12-13 | 07:45 | 7979 Cedar Dr     | Disorderly Conduct      |
| 2025-12-13 | 15:15 | 8080 Birch Ln     | Aggravated Assault      |
| 2025-12-14 | 11:00 | 8181 Spruce Pl    | Sexual Assault          |
| 2025-12-14 | 19:30 | 8282 Willow St    | Aggravated Battery      |
| 2025-12-15 | 08:15 | 8383 Ash Ave      | Disorderly Conduct      |
| 2025-12-15 | 16:45 | 8484 Hickory Rd   | Aggravated Assault      |
| 2025-12-16 | 10:30 | 8585 Sycamore Dr  | Sexual Assault          |
| 2025-12-16 | 18:00 | 8686 Dogwood Ln   | Aggravated Battery      |
| 2025-12-17 | 06:00 | 8787 Redwood Way  | Disorderly Conduct      |
| 2025-12-17 | 13:45 | 8888 Juniper Ct   | Aggravated Assault      |
| 2025-12-18 | 09:15 | 8989 Cypress St   | Sexual Assault          |
| 2025-12-18 | 17:45 | 9090 Fir Ave      | Aggravated Battery      |
| 2025-12-19 | 07:30 | 9191 Palm Rd      | Disorderly Conduct      |
| 2025-12-19 | 15:00 | 9292 Cedar Dr     | Aggravated Assault      |
| 2025-12-20 | 11:45 | 9393 Birch Ln     | Sexual Assault          |
| 2025-12-20 | 19:15 | 9494 Spruce Pl    | Aggravated Battery      |
| 2025-12-21 | 08:45 | 9595 Willow St    | Disorderly Conduct      |
| 2025-12-21 | 16:15 | 9696 Ash Ave      | Aggravated Assault      |
| 2025-12-22 | 10:00 | 9797 Hickory Rd   | Sexual Assault          |
| 2025-12-22 | 18:30 | 9898 Sycamore Dr  | Aggravated Battery      |
| 2025-12-23 | 06:30 | 9999 Dogwood Ln   | Disorderly Conduct      |
| 2025-12-23 | 14:00 | 10000 Redwood Way | Aggravated Assault      |

Within the data, key fields stand out for their role in examining patterns, forecasting outcomes - these include:

What kind of crime happened - like stealing, attacking someone, using illegal drugs, taking a car without permission, robbing with force, lying to get money, making fake documents, or similar acts

## Data Preprocessing With Feature Extraction



A look at how crime information gets ready for computer analysis shows a series

of stages. Each step shapes messy records into something usable. Cleaning happens first, removing errors that could mislead results later. Then sorting takes place, lining up entries by time or location

## CONCLUSION

A computer program was built here to study past crime records. It uses smart methods to guess how much crime might happen in different parts of a city. One part looks at old numbers, another checks spots where police work. Every corner of the city's safety needs gets weighed. With these facts, guesses come alive - what happened last year, which neighborhood, what kind of incident. From that, pictures form: how busy courts could get, when trouble rises, who might need help first.

When models run, they hit high marks - precision, recall, F1 - all topping out near 99.7 percent when guessing where, when, or who might be involved in crime. Outcomes show city zones light up more often, summer months spike, certain neighborhoods repeat. These visuals

help teams spot risky spots before trouble grows. With solid data guiding moves, police shift focus smarter, routes adjust, presence tightens where it counts most.

## REFERENCES

- Aggarwal, C. C.** (2015). *Data Mining: The Textbook*. Springer Publications.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al.** (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.
- McKinney, W.** (2010). *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference.
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G.** (2011). *The NumPy Array: A Structure for Efficient Numerical Computation*. Computing in Science & Engineering, 13(2), 22–30.
- Friedman, J. H.** (2001). *Greedy Function Approximation: A Gradient Boosting Machine*. Annals of Statistics, 29(5), 1189–1232.
- Breiman, L.** (2001). *Random Forests*. Machine Learning Journal, 45(1), 5–32.