

# MACHINE LEARNING FOR THE STUDY AND PREDICTION OF CRIMINAL DATA

<sup>1</sup>Mohammed Sadiq.B, <sup>2</sup>Sanjana C S

[1] Assistant Professor, Department of MCA, BIET, Davanagere

[2] Student, Department of Master of Computer Application, BIET, Davanagere

## ABSTRACT

*The goal of the Crime Data Analysis Project is to analyze and interpret enormous amounts of crime-related data in order to find insightful patterns and trends. In order to improve crime prevention methods, maximize resource allocation, and foster safer settings, this project aims to offer law enforcement agencies, lawmakers, and communities relevant information.*

*Gather pertinent crime data from a variety of sources, including law enforcement organizations, public records, and emergency services, while assuring data accuracy and consistency.*

**Keywords:**

*Criminology, Crime Analysis, Crime Prediction.*

## I. INTRODUCTION

The goal of the Crime Data Analysis Project is to investigate and examine crime-related data in order to identify insightful trends. This project aims to provide a thorough understanding of crime dynamics, detect patterns, and promote evidence-based decision-making for law enforcement agencies, lawmakers, and community stakeholders by analyzing diverse crime datasets.

Understanding crime's patterns and underlying causes is essential for developing effective crime prevention and intervention measures since crime is a major concern in contemporary society. Due

to the abundance of crime data, there is a chance to use data analysis methods and get practical knowledge that will help make communities safer.

This project's goal is to examine crime data using sophisticated data analysis and visualization approaches. We can find hidden patterns, spot high-risk areas, and gain understanding into what drives criminal activity by utilizing the power of data.

Key components of this project include:

- Data gathering: Compile statistics on crimes from dependable sources such law enforcement organizations, public records, and crime reporting systems. To achieve accurate analysis, make sure your data are reliable and consistent.
- Cleaning and preprocessing: the crime data is necessary to resolve problems including missing numbers, outliers, and discrepancies. Standardize the data's structure and format to enable efficient analysis.
- Exploratory Data Analysis: To fully comprehend the criminal data, use exploratory data analysis. To find patterns, correlations, and anomalies in the dataset, use statistical methods and visualizations.

### A. Problem statement

The goal of the Crime Data Analysis Project is to tackle the problem of efficiently utilizing data on crimes in order to obtain knowledge, spot trends,

and support decision-making procedures for crime prevention and intervention. The following is a definition of the project's problem statement:

**Inadequate use of crime data:** Due to a lack of effective methodologies and instruments for analysis, law enforcement organizations, lawmakers, and community stakeholders have trouble drawing conclusions from crime data that are useful. When it comes to finding hidden patterns, comprehending crime dynamics, and developing reliable predictions, traditional methodologies frequently fall short. The creation of efficient crime prevention measures and resource allocation are hampered by this restriction.

### ***B. Objectives***

**Data collection and integration:** assemble a range of criminal data from trustworthy sources such law enforcement organizations, public databases, and crime reporting systems. Ensure that data from many sources is included into a single analysis framework.

**Exploratory Data Analysis:** Investigate the crime data in detail to look for trends, correlations, and anomalies. Discover spatial and temporal trends in criminal activity using statistical techniques, data visualization, and geospatial analysis.

**Social media Analysis:** Using social media data and sentiment analysis techniques, you can learn about how the public feels about crime. Watch for new debates about crime on online forums, and assess how they affect crime rates.

## **II. EXISTING SYSTEM**

**Traditional statistical Analysis:** Traditional statistical analysis techniques are frequently used by law enforcement authorities to review crime data. Basic estimates of crime rates, trends, and correlations are often involved in this. These analyses frequently use statistical programs like

SPSS or spreadsheet programs like Microsoft Excel.

**Research-Based Systems:** For certain studies or initiatives, academic scholars and research institutions frequently create specialized crime data analysis systems. To investigate crime patterns, identify risk factors, and suggest evidence-based remedies, these systems may make use of cutting-edge statistical methods, machine learning algorithms, and data visualization tools.

## **III. PROPOSED SYSTEM**

The proposed system for the Crime Data Analysis Project intends to offer an advanced, thorough, and user-friendly platform for crime data analysis. The following elements and functionalities are included in the suggested system:

**Data Preprocessing and Integration:** The system will be able to gather and combine criminal data from a variety of sources, including public records, law enforcement databases, and social media platforms. To manage missing values, outliers, and inconsistencies, it will use strong data pretreatment techniques, ensuring data quality and suitability for analysis.

**Scalability and Flexibility:** The system will be made to be able to handle big crime datasets as well as adapt to changing data sources and analytic requirements. To provide a comprehensive understanding of crime trends, it will support the integration of various data types such as socioeconomic indicators, meteorological data, and demographic data.

**User-Friendly Interface:** With a user-friendly interface that enables non-technical users to navigate and utilize the features of the platform, the system will prioritize usability and ease of use. To support data-driven decision-making, it will include straightforward workflows, guided analysis phases, and interactive features.

#### IV. LITERATURE REVIEW

"Crime Pattern Detection Using Data Mining" by Johnson et al. (2018): In order to discover crime patterns, this study investigates the use of data mining techniques such as association rule mining and clustering. The authors show how using these tools can help with proactive crime prevention tactics by revealing hidden correlations and crime trends.

"Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations" by Mohler et al. (2017): The application of predictive modeling in criminal analysis is covered in this work. It emphasizes how crucial precise crime forecasting is for streamlining resource allocation and enhancing police deployment tactics. The study provides a thorough analysis of different predictive policing techniques and their efficacy.

"Crime Forecasting Using Machine Learning Techniques" by Bhatt et al. (2019): The application of machine learning methods, such as random forests and support vector machines, for crime prediction is examined in this paper. The authors address the potential of machine learning in raising the accuracy of crime predicting while comparing the results of several models.

"Social Media Analysis for Crime Prevention and Prediction: A Survey" by Shrestha et al. (2020): This review article examines how social media analysis is used to predict and prevent crime. It investigates tools like sentiment analysis, social network analysis, and text mining to gather data on crimes from social media platforms. The paper explores the advantages and disadvantages of using social media data for criminal investigation.

#### V. IMPLEMENTATION

Algorithm used,

Two algorithms are used KNN and Decision Tree, both are Supervised

**KNN(K-Nearest Neighbor):** One of the popular machine learning techniques that can be used to

analyze crime data is the K-Nearest Neighbors (KNN) algorithm. KNN is a classification technique that works well with datasets that have instances that have labels.

**Decision Tree:** Using a Decision Tree algorithm to analyze crime data can be an effective way to spot trends, pinpoint key variables, and forecast crime outcomes. Popular machine learning techniques known as decision trees are capable of performing both classification and regression tasks. We will concentrate on classification as it relates to the study of crime data.

#### VII. METHODOLOGY

A systematic methodology is frequently used in crime data analysis projects to ensure that the study is carried out quickly and successfully. The following is a step-by-step process for a project analyzing crime data:

**Define the Objective:** Clearly state the goal and aim of the project analyzing crime data. Make a list of the precise queries you want to address or issues you wish to solve through analysis. For instance, you might want to pinpoint crime hotspots, comprehend historical crime trends, or assess the efficiency of crime prevention measures.

**Identify Data Sources:** Find out where the crime statistics that will be utilized in the analysis came from. Police reports, incident reports, victimization surveys, court statistics, and any other pertinent sources may be used in this. Recognize the reliability, accessibility, and restrictions of the data sources.

**Data Gathering and Cleaning:** Gather the relevant crime statistics from the listed sources. The data is cleaned by eliminating duplicates, addressing missing information, and fixing any mistakes or discrepancies. Make sure the data is in an analysis-ready format.

## V. MODULES DESIGN

There are different operational module in this application.

### Model 1 or baseline:

The training data's dataset of various city police stations is all that is predicted. On any feature, there is no bias.

### Model 2:

I made an effort to incorporate the Location, Crime Type, Crime Nature, and Crime Year into this model. The code I used was from pandas. My file reads and computations became incredibly quick as a result. Pandas uses C to carry out the calculations, whereas Python is only used as a wrapper. I separate the training data for this model based on these three features.

### Model 3:

According to the data description and the name of the company, the criminal category I used for this model is not particularly broad. I created a subset of data using the group by feature in pandas, where each row reflects a different average combination of the category and crime categories.

### A. System Architecture

The Crime Data Analysis Project's system architecture includes all of the necessary elements and modules for efficient data processing, analysis, and display.

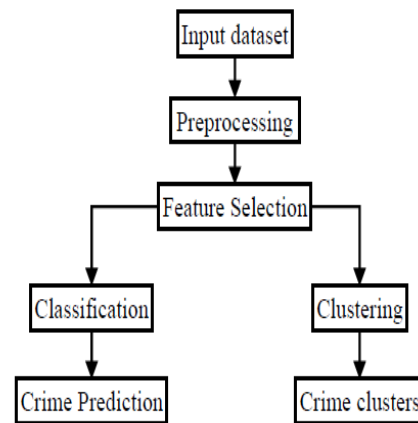


Fig. 1 System Architecture

## VI. RESULTS

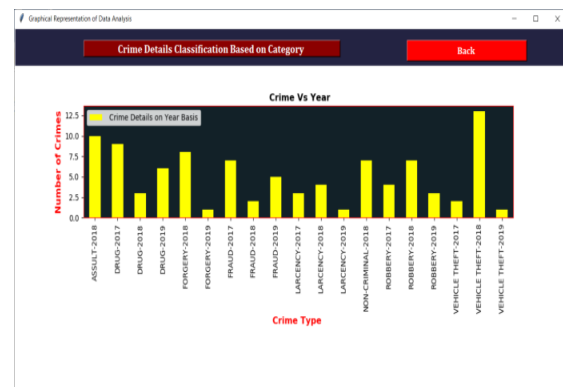
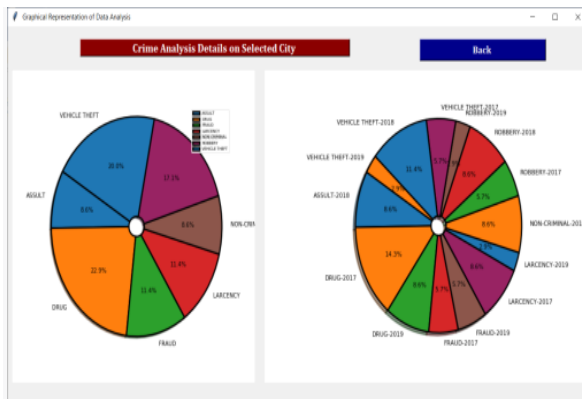


Fig: Analyzed City Crime Data based on Year.



Fig: Analyzed City Crime Data based on Year.



**Fig. 12 Graphical Representation of Analyzed Data**

## VII. CONCLUSION

In this application, we developed and put into use a machine learning software model to process a dataset of criminal data past and forecast citywide crime rates. A software model called Criminal Data Analysis and Prediction uses all the requirements for municipal crimes to calculate efficiently and accurately the criminal details of a city with individual police department stations. Given the information, we can forecast the history of the crime rate in a specific location, year, and kind. We may design a number of scenarios, such as the case rate in a city, using this knowledge.

## REFERENCES

- [1] Muddha Sharma, "Z-Crime: A Data Mining Tool for the Detection of Suspicious Criminal Activities based on the Decision Tree", International Conference on Data Mining and Intelligent Computing, pp.1-6, 2014.
- [2] Kaumalee Bogahawatte and Shalinda Adikari, "Intelligent Criminal Identification System", Proceeding of 8<sup>th</sup> IEEE International Conference on Computer Science and Education, pp.633-638, 2013.

- [3] jyoti Agarwal, Renuka Nagpal and Rajni Sehgal, "Crime Analysis using K-Means Clustering", International of Journal Computer Applications, Vol.83, No.4, pp.1-4, 2013.

- [4] Rasoul Kiani Siamak Mahdavi and Amin Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification", International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No. 8, pp. 11-17, 2015.

- [5] Shiju Sathyadeyan, M.S. Devan and S. Surya Gangadharan, "Crime Analysis and Prediction using Data Mining", Proceedings of IEEE 1<sup>st</sup> International Conference on Networks and Soft Computing, pp. 406-412, 2014.