



## Crime Rate Prediction Using Machine Learning

*S. Ayesha Tanveer<sup>1</sup>, Bonthala Khyathi<sup>2</sup>, Vardhamsetty Guru Sri Latha<sup>3</sup>, Thalla Jagadeeswar<sup>4</sup>*

Department of ECE, Madanapalle Institute of Technology & Science, Madanapalle, India

### Abstract

In India, there are continuously more criminal cases filed, which results in an increase in the number of cases still outstanding. These ongoing increases in criminal cases make them challenging to categorise and resolve. Therefore, it's crucial to identify a location's patterns of criminal activity in order to stop it from happening in order to lessen crime. If crime-solving organisations have a clear understanding of the trends in criminal activity in a certain area, they will be able to perform their function more effectively. These can be carried out utilising machine learning and a variety of methods.

Here we are using Random forest classifier algorithm to find the patterns of the criminal activities in a particular area. This project uses data set and predicts the type of crime in a particular area which helps in speeding up the classification of criminal cases and proceed accordingly. Data preprocessing is as important as final prediction, this project used feature selection, removing null values and label encoding to clean and nourish the data. This research gives an efficient machine learning model for predicting the next criminal case.

### INTRODUCTION

An in-depth investigation and analysis should be carried out within to resolve a case based on specific facts. The analysis and decision-making of these offenders has never been simpler because to the volume of crime data currently available in India. The cases are too difficult for the officials to handle. This program which focuses on finding a solution for crime decision-making, identifies this as a significant problem.

This research's main objective is to forecast criminal activity using a pre-organized data set in order to reduce crime rates. The fundamental problem is that when populations increase, crime rates increase as well, making it hard for authorities to predict crime rates with any degree of accuracy. Because they are concentrating on so many different issues, the authorities might not be able to predict upcoming crimes. The government and police personnel cannot promise a 100% reduction in crime, despite their best efforts. They may find it difficult to forecast the level of crime in the future. Numerous studies have been conducted in connection with crimes. Large datasets have been analysed, and data like location and crime type has been extracted to help follow police enforcement. Existing methods have made use of these databases to find and identify crime hotspots.

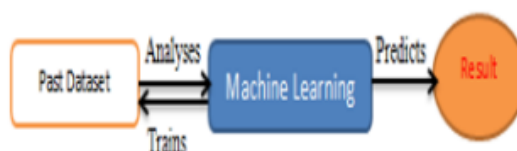


Fig-1: Process of Machine Learning

### RELATED WORK

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is secondary sources and discuss published information in a particular subject area and sometimes information in a particular subject area within a certain time period.

Its ultimate objective is to update the reader on the state of a particular field of literature and serves as the foundation for additional objectives, future studies that may be required in the field, for example, prior to a research proposal, which could just be a brief synopsis of the sources. It often follows an organisational structure.

It incorporates both synthesis and summation. Previous studies examined if there was a positive linear link between penalties and violent crimes after disproving the widespread belief that unemployment and violent crimes are substantially correlated. We found no association at the local level because of the complicated ways that boundaries are established and offences are classified. However, the linear association became evident and statistically significant at the state level. Overlaps between the top states for fines and the top states for violent crime verified the results of our fit. Additionally, it covered how police handle different demographic subgroups fairly when conducting stops and searches.

The question of violent crime's causes is extremely complex.

It demonstrated a connection between places with high fines and high rates of violent crime. It studies and evaluates the dependence of city and county revenue derived from fines (mainly traffic violations) and their potential implications on the incidence of violent crimes on an aggregated state level. Crime exists, and there are potential consequences of excessive fining in some regions. Following the riots, multiple press reports discussed how Ferguson's high municipal court fines (again, typically for traffic offences) undermined locals' confidence in the police and the city government as a whole. It examined if the practise of generating sizable municipal income from minor infractions has an effect on violent crimes not only in Missouri but also in other states.

Around one billion people are thought to live in India. The high population density will increase the rate of violence when paired with other factors like unemployment, poverty, and illiteracy.

## OVERVIEW OF SYSTEM

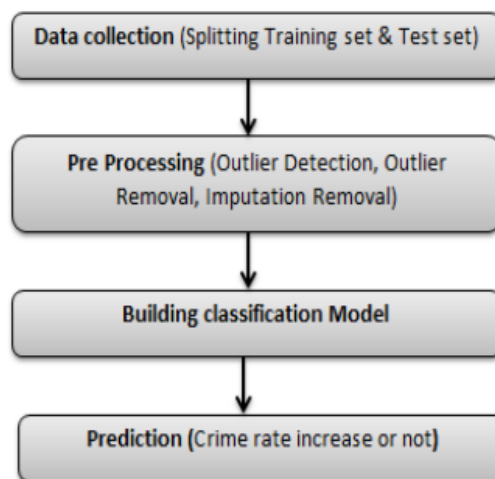
Testing a model's ability to anticipate would be the goal. Utilizing the training dataset, the test dataset will be used to validate the training. Depending on the accuracy, a better algorithm will be used to build the model. For crime prediction, supervised classification and other algorithms will be employed. To examine potential crimes that may have occurred in the nation, the dataset is visualised. The crime rate in India is decreased as a result of this work's assistance to law enforcement organisations in predicting and detecting crimes more accurately. This facilitates the completion of other formalities by all other departments.

Building a model that can make predictions is done through predictive modelling. A machine learning algorithm that learns specific attributes is used in the process.

As the result of a prediction, a classification model assigns discrete class labels to a specific data value. A pattern classification problem in weather forecasting could be the prediction of a sunny, wet, or snowy day. This is an example of a classification model. The two categories of supervised learning and unsupervised learning can be used to split pattern classification jobs. In supervised learning, the classification model's input dataset's class labels are predetermined. When dealing with a supervised learning problem, it would be aware of which training dataset has the specific output that will be trained so that prediction might be made. This analysis looks at general tendencies that could aid in model and hyper parameter selection as well as which features are most useful in The discrepancy between actual records and our predicted values for both years indicates that the suggested approaches are between 85% and 90% accurate.

In the future, when there are enough data points to use machine learning (ML) models, this study can be modified to forecast crime using ML models. Additionally, it may improve prediction accuracy. Additionally, statistical modelling techniques can be combined with ML models to generate weighted accuracy for a district, which helps strengthen the solution. predicting crime rate.

The objective is to categorise the level of criminality in order to find fraud hotspots in the actual world.



**Fig-2: Data Flow Diagram for Machine Learning Model**

## EXISTING SYSTEM

Modern technological advancements in sophisticated data analytics and visualisation tools are assisting society in various ways to examine the data that is relevant to social issues. Crime statistics for various demographic areas are one of these socially relevant activities. The examination of crime data will aid decision-making organisations in taking preventative measures to reduce crime in populous areas.

Publicly accessible information and services, along with advancements in information technology, somehow assist criminals in carrying out their misdeeds and entangling them in more serious crimes than before. As a result, both industrialised and developing countries are experiencing extraordinarily rapid increases in crime. It presents statistical models using weighted moving average and functional statistics based on the previous year's crime data for Indian states.

---

## PROPOSED SYSTEM

Users have access to a wide variety of machine learning techniques that can be used with datasets. However, supervised learning and unsupervised learning algorithms are the two main categories of learning algorithms. The "correct answer" is inferred by supervised learning algorithms using labelled training data. A specific attribute or group of qualities is provided to the algorithms to forecast. Methods for removing null values and infinite values that could impair the system's accuracy are part of the data preparation process. Formatting, cleaning, and sampling are the primary procedures. There may be incomplete data that has to be fixed or removed using the cleaning process.

Here we are using Random Forest Classifier algorithm. To predict the crime rate of illegal activities, we are utilising the random forest classifier method in the suggested system. Here, we are using real-time data that we have gathered from the nearby police station to analyse the pre-processing data in an effort to lower crime rates. The research that was conducted by looking through many such documentations served as the foundation for the suggested system.

Almost all crimes are predicted based on where they occur and the kinds of crimes that are most common there. According to the state and year, the data set includes various crime kinds that are perpetrated in India. The area where crimes are committed is the output of this article, which uses crime kinds as its input. Data cleaning, feature selection, removing null values, and data scaling via normalising and standardising are all part of the data pre-processing process. Following data preparation, the data is free of null values that could dramatically influence the model's accuracy, and feature selection is used to choose only the necessary characteristics that won't affect the model's accuracy.

The chosen models—Logistic Regression, Decision Tree, and Random Forest—are then trained by dividing the data into train and test sets after data pre-processing. Classification models are utilised in this situation since the output that is needed is a category value. The data prediction is done using the Python language. With this algorithm for a data mining strategy to help forecast the patterns of crime and expedite the process of solving crime, we had a high degree of accuracy in our model prediction methodology. Using this technology, the results are precise and faultless.

---

## DATA COLLECTION

Here we have collected the dataset from our near by police station which contains the information like type of crime, section, name of accused person, name of witness, age, gender, report date, report time, occurrence date and occurrence time etc. The Data Model which was created using Random Forest, Decision tree algorithms are applied on the Training set and based on the test result accuracy, Test set prediction is done.

### Data Processing

This procedure contains steps to eliminate any null or infinite values that can impair the system's accuracy. Formatting, cleaning, and sampling are the primary procedures. There may be incomplete data that has to be fixed or removed using the cleaning process. When adequate data are utilised in sampling, the algorithm's execution time may be shortened. The preprocessing is carried out in Python. The data that was gathered can have missing values, which could cause it to be inconsistent. Data must be preprocessed in order to increase the algorithm's effectiveness and produce better results.

The outliers must be eliminated, and variable conversion must be performed. Considering the relationship between the traits, it was observed that property area, education, loan amount, and finally credit history—which is the most important of all—are characteristics that are noteworthy on their own. Contrary to common belief, several variables, such as applicant and co-applicant income, are not significant when analysed separately. The data is condensed to a set minimum of records to facilitate analysis. The Indian police agency updates and maintains the dataset that was collected online.

Development of a Predictive Model Data collection is necessary for machine learning, and there is a wealth of historical data. There is enough historical data and unprocessed data for data collection. Raw data cannot be used directly without preparation. The kind of algorithm employed with the model is then preprocessed. This model has been tested and trained, and it makes accurate predictions with little errors. A tuned model is continually adjusted to increase accuracy. Finally, after the model has been developed, tested, and deployed, perform an analysis of the provided dataset and explain how to automatically fix it in light of the inconsistent historical data on bank accountants.

---

## TESTING THE DATASET

Steps in testing the dataset, are the following

1. The predict method, which accepts this array as input and produces the expected target value as output, is now used to forecast the species of the new features in a numpy array called "n."
2. As a result, the expected goal value is zero. By comparing the actual values of the test set with the projected values, it is possible to determine the test score, which is calculated as the ratio of the number of correctly predicted outcomes to the total number of predictions made.
3. The first line imports the preconfigured iris data set from the Sklearn module. The raw data set is essentially a table that provides details about distinct types.
4. For use in this application, import any algorithm and the train test split class from the sklearn and numpy modules.
5. To include the data dataset variable's load data() method. Using the train test split function, further split the dataset into training and test data. The variable's X and Y prefixes stand for the feature and target values, respectively.
6. Using a ratio of 67:33:70:30, this method randomly divides the dataset into training and test data. Then any algorithm is enclosed.
7. In order to train the computer, we fit our training data into this algorithm in the following line.

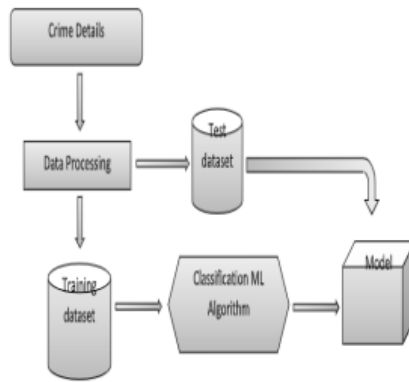


Fig-3: Architecture of Proposed Model

CONCLUSION

Finding relationships and patterns among varied data has become much simpler with the use of machine learning technology. The major task of this research is to determine the type of crime that might occur given the place where it has already happened. Using a training set of data that has undergone data cleansing and data transformation, we have developed a model using the machine learning idea. With an accuracy of 0.789, the model can identify the type of crime. Analyzing a data set is made easier by data visualisation. The graphs include bar, pie, line, and scatter diagrams, each with their unique features. We created numerous graphs and discovered intriguing data that assisted in studying crime datasets that can assist in identifying the elements that can help in keeping society clean.

RESULTS

```

In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import math

In [2]: df=pd.read_csv("C:/Users/dell/Desktop/crimerate/New Data.csv")
df.head()

Out[2]:
   s.no  town section  type of crime  report date  Report Time  Occurance Time  Occurance Time  occurrence date  name of the accused persons  age of accused person  gender of accused person  address
0  1.0  madanapalli  UH  women missing  12-30-2021  11:00 PM  02:30 PM  1-1-2022  ...  ranamma  33.0  female  k...
1  2.0  madanapalli  UH  women missing  1-2-2022  03:30 PM  12:45 PM  1-2-2022  ...  subramanyan & son others  31.0  male  va...
2  3.0  madanapalli  337.279 IPC  accident case  1-2-2022  12:30 AM  04:30 PM  1-3-2021  ...  dana naddy  35.0  male  mada...
3  4.0  madanapalli  324 Iw 34 IPC  violence  1-3-2022  11:45 PM  04:45 PM  1-1-2022  ...  rahamathulla  50.0  male  mada...
4  5.0  madanapalli  379 IPC  robbery  1-4-2022  04:30 PM  06:30 PM  1-2-2022  ...  banu prakash  24.0  male  mada...

5 rows x 25 columns

In [3]: df.isna().any()
df.shape

Out[3]: (260, 25)

In [4]: df=df.dropna()
df

Out[4]:
   s.no  town section  type of crime  report date  Report Time  Occurance Time  Occurance Time  occurrence date  name of the accused persons  age of accused person  gender of accused person  address
0  1.0  madanapalli  UH  women missing  12-30-2021  11:00 PM  02:30 PM  1-1-2022  ...  ranamma  33.0  female  female
    
```

Fig-4:Result1

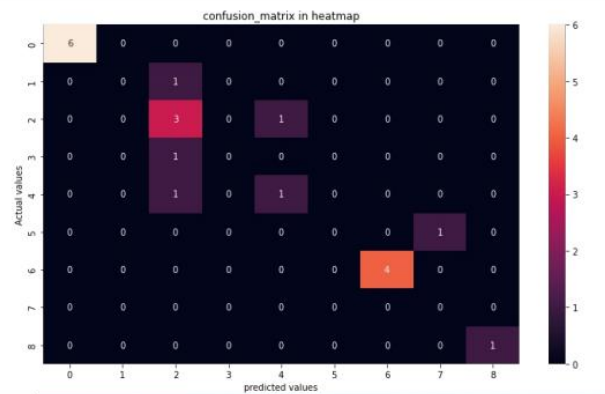


Fig-5: Result 2

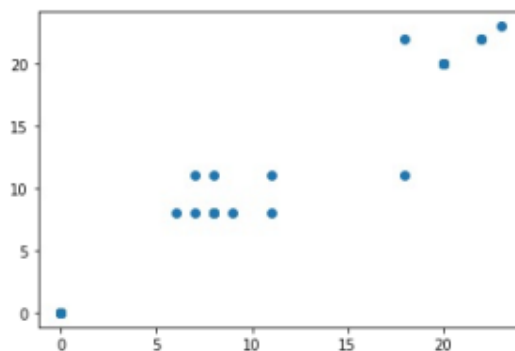


Fig- 6: Result 3

## REFERENCES

1. Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of (Vol. 1, pp. 225- 230). IEEE.
2. Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.
3. Sivaranjani, S., Sivakumari, S., & Aasha, M. (2016, October). Crime prediction and forecasting in Tamilnadu using clustering approaches. In Emerging Technological Trends (ICETT),
4. Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In Networks & Soft Computing (ICNSC), 2014 First International Conference on (pp. 406-412). IEEE.
5. Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 IEEE/WIC/ACM International Conference on (pp. 41-44). IEEE.
6. Zhao, X., & Tang, J. (2017, November). Exploring Transfer Learning for Crime Prediction. In Data Mining Workshops (ICDMW), 2017 IEEE International Conference on (pp. 1158-1159). IEEE.
7. Al Boni, M., & Gerber, M. S. (2016, December). Area Specific Crime Prediction Models. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 671-676). IEEE.
8. Tayebi, M. A., Gla, U., & Brantingham, P. L. (2015, May). Learning where to inspect: location learning for crime prediction. In Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on (pp. 25-30). IEEE.