iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** X **Month of publication:** October 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.64542

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue X Oct 2024- Available at www.ijraset.com*

# Crime Type and Occurrence Prediction Using Machine Learning Algorithm

Sajitha P[1], Dr. Arul Selvan A[2]
[1]*M.Sc. Software Systems, KG College of Arts and Science, Coimbatore, Tamil Nadu, India*
[2]*Assistant Professor MCA., M.Phil.,B.Ed.,Ph.D., PG Department of Software Systems and Computer Science, Bharathiar University, KG College of Arts and Science, Coimbatore, Tamil Nadu, India*

***Abstract:** This project is entitled as "Crime Type and Occurrence Prediction Using Machine Learning Algorithm" Crime is still a major worry and a serious problem in our society. As a result, crime prevention is a significant issue that needs to be examined methodically. Detecting and preventing crime requires effective crime analytics, which is also crucial for assessing how well criminal investigations are working. To create precise forecasts, inferences are drawn from trained data. a system that uses user input to forecast different characteristics of crime. Users select the year, offense type, and city name. The method forecasts the population, expected number of occurrences, and expected crime rate for the year in 100,000 units based on this data. The dataset used in this investigation was produced on its own. The system makes use of machine learning techniques, particularly Random Forest, SVM, Logistic Regression, and Linear Regression. To guarantee precise forecasts and dependable results, it has undergone testing and training.*

## I. INTRODUCTION

A crime is an act that qualifies as an offense. It is governed by the law. For the police, identifying and analyzing hidden crimes is an extremely challenging task. Furthermore, there is a wealth of criminal data available, therefore methods are required to make investigation easier. Consequently, the approach ought to be helpful in resolving criminal cases. Crime analysis and prediction can be further aided by machine learning techniques. Regression techniques are provided by the machine learning methodology. The classification methods aid in achieving the goal of the inquiry. Regression techniques like logistic regression and linear regression.This technique aids in determining how two quantitative values or variables relate to one another. This method uses the independent variables to predict the value of the dependent variable. classifier techniques like Random Forest, SVM, etc. Multi-class target variables are classified using these classifiers. Accuracy is increased by using a neural network. The neural network consists of an output layer and a dense input layer. The perpetrator's description, including prediction rate, estimated crime rate, estimated number of cases, and population in lakhs, is based on the algorithms mentioned above. As a result, she will assist in solving murder cases, which should lessen the load on police investigations.

## II. OBJECTIVES OF THE PROJECT

1) To increase law enforcement's capacity to proactively prevent and reduce criminal activity, machine learning techniques are being used to forecast the types and incidences of crimes. The objective is to accurately anticipate the kind of crime that is likely to occur in a specific location and at a specific time by utilizing machine learning algorithms and examining current crime data.
2) By detecting crime hotspots and preventing crimes before they happen, this predictive power can help government better allocate resources and lower crime rates.The objective is to use technology to strengthen public safety and crime prevention tactics.

## III. STUDY ON EXISTING SYSTEM

Models for predicting crime frequently use historical data, which may have built-in biases based on geography, socioeconomic position, or race. For minorities in particular, this might result in skewed projections and sustain structural injustices. Incomplete or noisy data may decrease the accuracy of crime prediction models. Predictions may not be accurate since crime is influenced by intricate social and psychological elements that are hard to record in data. Confidential data is frequently processed by crime predicting systems, making them possible targets for cyberattacks. Both public security and individual confidentiality may be jeopardized by violations of these systems.

©IJRASET: All Rights are Reserved 

### A. *Drawbacks*

1) The classifier's use of a categorical value biases the output for nominal qualities with higher values, contributing to low accuracy in previous efforts.
2) The classification technique is not suitable for regions with inaccurate data and real-valued attributes
3) To optimize the classifier's performance, its value must be modified
4) Identifying crime patterns and extracting information from big amounts of data is comprehensive.
5) Many crime prediction systems involve complicated algorithms, also known as "black-box" models, that are difficult to explain.
6) The incorporation of personal data in crime prediction models, such as social media activity, surveillance footage, or mobile phone data, poses substantial privacy concerns.

## IV. PROPOSED SYSTEM

A machine learning model that predicts the complex model and trend of a large dataset. This can identify crimes across geographies and acts. This leads to a more accurate prediction of crime in a historical database, including time, location, and kind of crime. Analyzing behavioral patterns and suspicious conduct can aid in the early detection and prevention of criminal activity. For example, ML models can be used to anticipate specific crimes, such as robbery or fraud, based on predetermined triggers. Predictive data enables law enforcement organizations to better allocate resources to high-risk regions, thereby lowering crime and increasing public safety.

### A. *Advantages*

1) Optimal values do not need to be initialized.
2) High accuracy compared to other machine learning prediction models.
3) Eliminates the need to analyze independent attributes.
4) The suggested method accurately detects criminal patterns based on prediction rate, anticipated crime rate, and number of instances. It also eliminates the need to analyze independent effects.

## V. METHODOLOGY

1) Data collection
2) Data preprocessing
3) Analysis
4) Training and Testing
5) Validation

### A. *Data Collection*

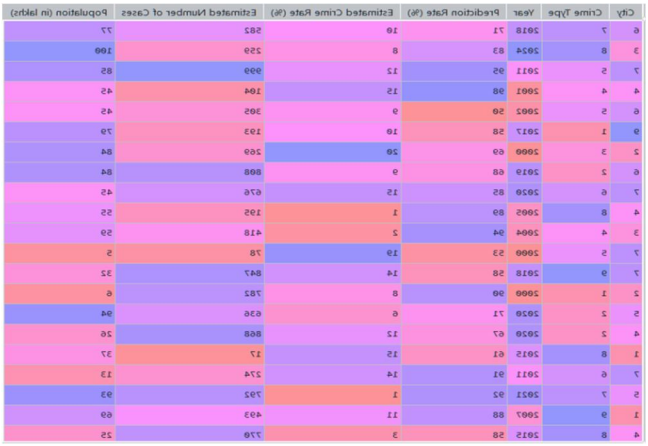
The data for the implementation is sourced from Kaggle.The record collection totals over 3000.This dataset should preferably include information such as city, type, year, population, predicted rate, and estimated rate.

| | City | Crime Type | Year | Prediction Rate (%) | Estimated Crime Rate (%) | Estimated Number of Cases | Population (in lakhs) |
|---|---|---|---|---|---|---|---|
| 2 | 6 | 7 | 2018 | 71 | 10 | 582 | 77 |
| 3 | 3 | 8 | 2024 | 83 | 8 | 259 | 100 |
| 4 | 7 | 5 | 2011 | 95 | 12 | 999 | 85 |
| 5 | 4 | 4 | 2001 | 98 | 15 | 104 | 45 |
| 6 | 6 | 5 | 2002 | 50 | 9 | 305 | 45 |
| 7 | 9 | 1 | 2017 | 58 | 10 | 193 | 79 |
| 8 | 2 | 3 | 2000 | 69 | 20 | 269 | 84 |
| 9 | 6 | 2 | 2019 | 68 | 9 | 808 | 84 |
| 10 | 7 | 6 | 2020 | 85 | 15 | 676 | 45 |
| 11 | 4 | 8 | 2005 | 89 | 1 | 195 | 55 |
| 12 | 3 | 4 | 2004 | 94 | 2 | 418 | 59 |
| 13 | 7 | 5 | 2000 | 53 | 19 | 78 | 5 |
| 14 | 7 | 9 | 2018 | 58 | 14 | 847 | 32 |
| 15 | 2 | 1 | 2000 | 90 | 8 | 782 | 6 |
| 16 | 5 | 2 | 2020 | 71 | 6 | 636 | 94 |
| 17 | 4 | 2 | 2020 | 67 | 12 | 868 | 26 |

©IJRASET: All Rights are Reserved 

### *B. Data preprocessing*

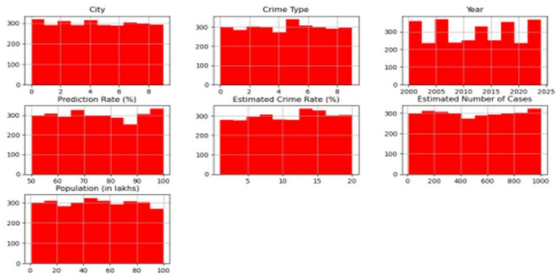
The dataset has been collected; however, it needs be preprocessed in order to provide a clean dataset. Python includes libraries such as pandas and numpuy.One-hot encoding techniques can be used to encode categorical variables (such as city, crime type, and year) into numerical representation.



### *C. Analysis*

The analysis contains a graphical depiction of several variables used to analyze the dataset's properties. The various graphs are created using Matplotlib packages.Graphs include histogram plots, heatmaps, and crime type count graphs.

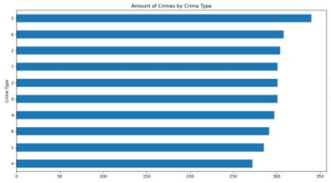
Histogram plot



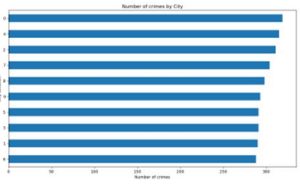

Heatmap



Crime type count graph

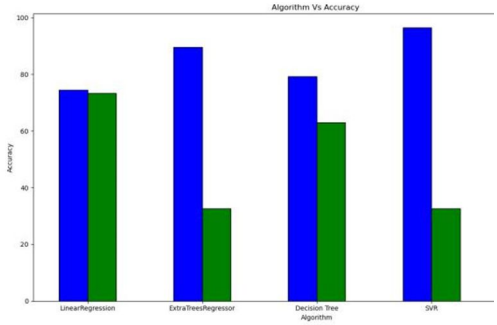


Crime city count graph



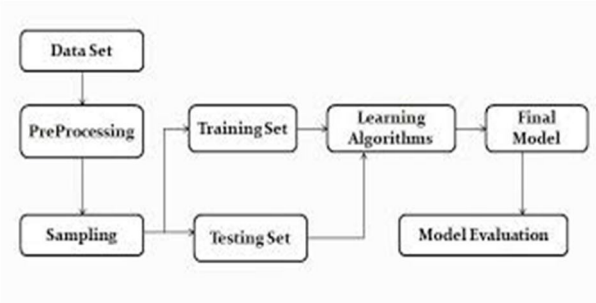
*D. Training and Testing*

The dataset is separated into two parts: training and testing. In general, 80% of the dataset is preserved for training, and 20% for testing. This divide allows the model to be assessed on previously unseen data.




*E. Evaluation*

After the model is constructed, it should be evaluated against real-time data values. This process is referred to as validation. The validation is nothing but the projected value, also known as the output value.




## VI. ALGORITHMS AND DESCRIPTION:

*A. Linear Regression*

One method of supervised learning is linear regression. Based on a given independent variable (X), it is in charge of forecasting the value of a dependent variable (Y). The relationship between the input (X) and the output (Y) is what it is. It is among the most well-known and comprehended algorithms for machine learning. The linear regression models are Regularization, Gradient Descent, Ordinary Least Squares, and Simple Linear Regression.

*B. Logistic Regression*

A supervised machine learning approach called logistic regression is used for classification problems in which predicting the likelihood that an instance belongs to a particular class or not is the aim. A statistical procedure called logistic regression examines the connection between two data points. The basics of logistic regression are examined in this article.

*C. SVM*

By carrying out optimal data transformations that establish boundaries between data points based on predefined classes, labels, or outputs, supervised learning models enable support vector machines (SVMs), a type of machine learning algorithm, to solve challenging classification, regression, and outlier detection problems. SVMs are extensively used in a variety of sectors, including speech and picture identification, natural language processing, healthcare, and signal processing applications.

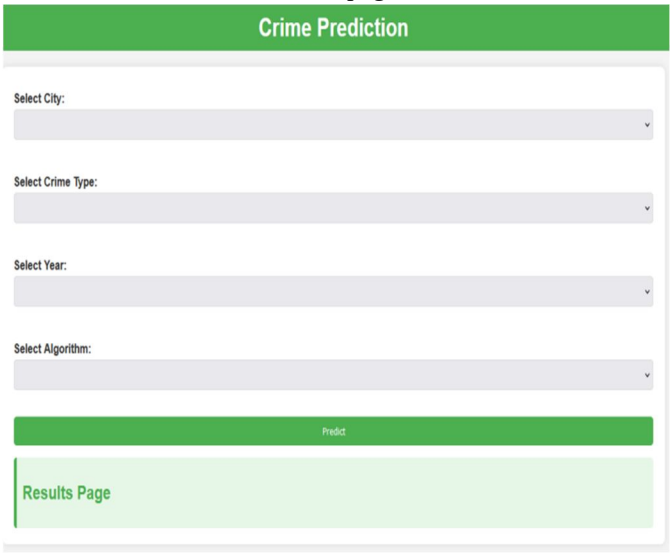
*D. Random Forest*

One well-known machine learning method that is a part of the supervised learning approach is Random Forest. In machine learning, it can be applied to both classification and regression issues. It is predicated on the idea of ensemble learning, which is the act of merging several classifiers to solve a challenging issue and enhance the model's functionality. Compared to previous algorithms, it requires less training time; it predicts output with high accuracy, especially for huge datasets; and it can retain accuracy even when a significant amount of the data is missing.

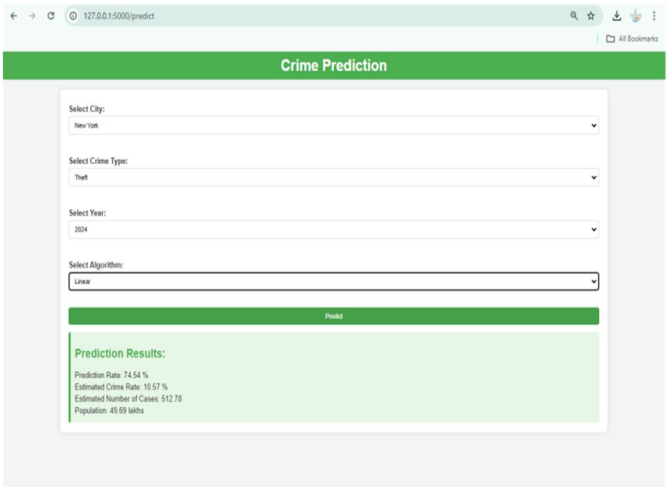
©IJRASET: All Rights are Reserved 

*E. Result*

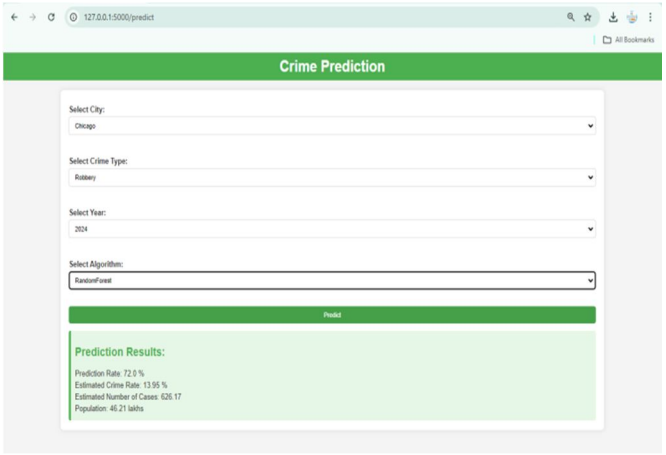
Home page




Linear regression




Random forest



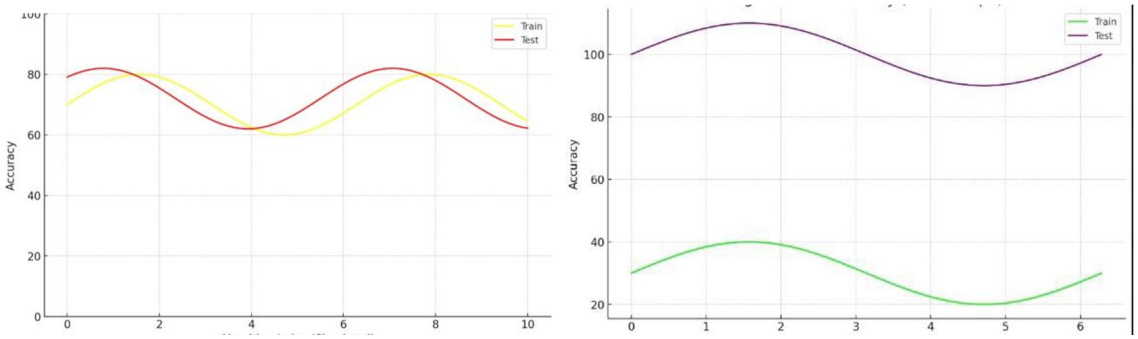
Logistic



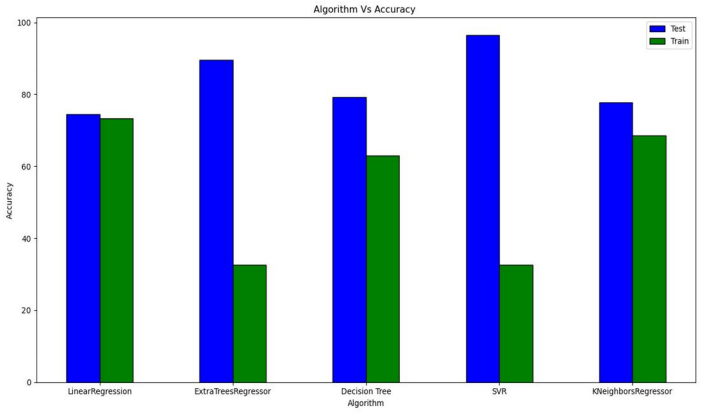

SVM



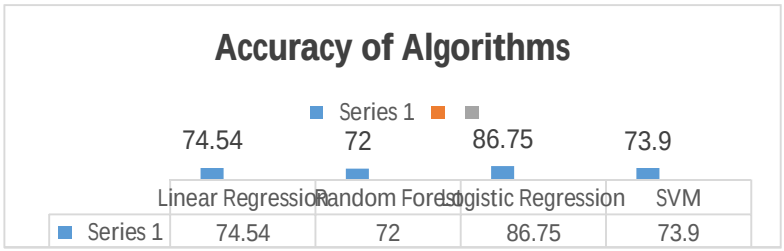

*F. Accuracy Graph*




*G. Testing and Training*



*H. Comparison of Algorithms*




## VII. FUTURE WORK

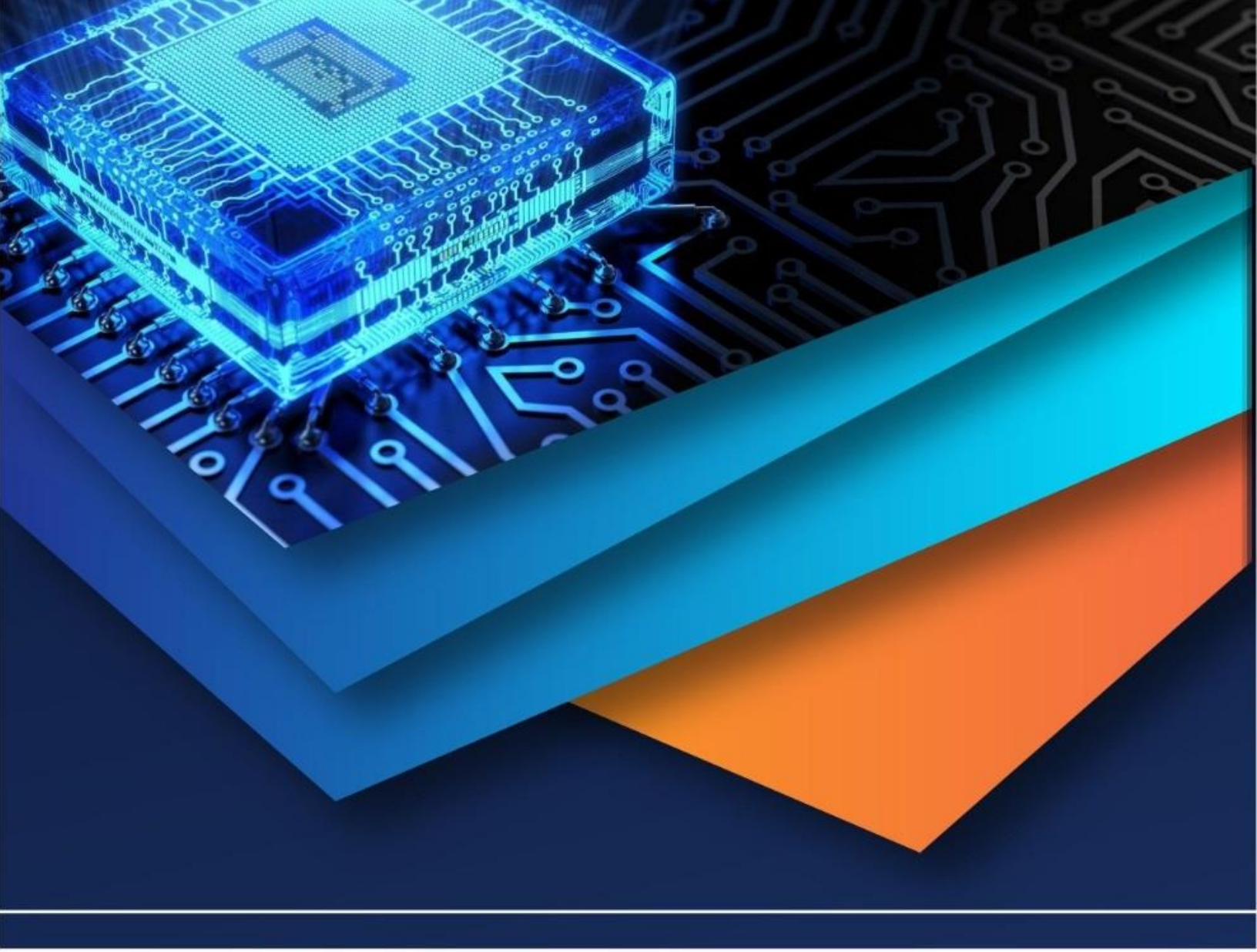
In order to enhance overall performance and increase the accuracy of crime prediction, future research in this field should concentrate on using additional categorization models. Comparing and contrasting the accuracy of the different classification algorithms for forecasting the kind of crimes that would occur would be beneficial. By examining the relationship between neighborhood income level and crime rate, it may be possible to gain a better understanding of the root causes of crime and create focused strategies to prevent it in high-risk regions. This kind of study may be helpful to governments and law enforcement agencies in better allocating resources to reduce crime and improve safety.

## VIII. CONCLUSION

Consequently, the machine learning model that makes use of SVM, random forest, linear regression, and logistical regression is highly effective in predicting criminal activity. India's crime rate rises daily due to a number of causes, including corruption, poverty, and poor execution. When taking the required actions to lower crime, the suggested model is very helpful to both police officials and investigating agencies. Through a variety of interactive visualizations, the project assists the criminal analysis in analyzing these crime networks. Future research in this field will focus on training robots to use automated learning techniques to forecast crime in a region. Since machine driving and data extraction are similar, autonomous learning expansion concepts can be utilized to improve predictions.To make better forecasts, increase the accuracy, dependability, and privacy of your data.

## REFERENCES

[1] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
[2] Machine Learning in Production: Developing and Optimizing Data Science Workflows and Applications (Addition-Wesley Data & Analytics)
[3] A.U.S. S Pradeep, SorenGoyal, J. A. Bloom, I. J. Cox, and M. Miller, ―Detection of statistical arbitrage using machine learning techniques in Indian Stock market, I IIT Kanpur, April 15, 2013.
[4] Prashant S. Chavan, Prof. Dr. Shrishail. T. Patil ―Parameters for Stock Market Prediction,I Prashant S Chavan et al, Int.J.Computer Technology & Applications, Vol 4 (2),337-340.
[5] NeelimaBudhani, Dr. C. K. Jha, Sandeep K. Budhani―Prediction of Stock Market Using Artificial Neural Network,I International Conference on Soft Computing Techniques for Engineering and Technology (ICSCTET)- 2014.
[6] SharvilKatariya, Saurabh Jain―Stock Price Trend Forecasting using Supervised Learning Methods.
[7] Chen, Ling, and Xu Lai. "Comparison between ARIMA and ANN models used in short-term wind speed forecasting." Power and Energy Engineering Conference (APPEEC), 2011 Asia- Pacific. IEEE, 2011.
[8] Agarwal, Jyoti, Renuka Nagpal, and Rajni Sehgal. "Crime analysis using K-means clustering." International Journal of Computer Applications 83.4 (2013).
[9] Sathyadevan, Shiju, and Surya Gangadharan. "Crime analysis and prediction using data mining." Networks & Soft Computing (ICNSC), 2014 First International Conference on. IEEE, 2014
[10] McClendon, Lawrence, and Natarajan Meghanathan. "Using machine learning algorithms to analyse crime data." Machine Learning and Applications: An International Journal (MLAIJ) 2.1 (2015).
[11] Kiani, Rasoul, Siamak Mahdavi, and Amin Keshavarzi. "Analysis and prediction of crimes by clustering and classification." Analysis 4.8 (2015)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)