e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

REVIEW ON CRIME ANALYSIS AND PREDICTION

Twinkle Antony Ayyanthole*1, Ms. Resija P.R*2

*1Department Of Computer Science, Vimala College (Autonomous) Thrissur, Kerala, India.

*2Assistant Professor, Department Of Computer Science Vimala College (Autonomous) Thrissur, Kerala, India.

DOI : https://www.doi.org/10.56726/IRJMETS45728

ABSTRACT

It is noticeable that crime rates [1] are rising daily throughout all the societies in the world. Data science is essential for predicting crimes. It helps us to find the patterns, correlations and trends by analyzing enormous amounts of crime-related data. In this review, we study the various algorithms, methodologies and models used for the crime analysis [2]. The main aim of this review is to study the prediction of crime analysis, hot spots and trends. Organizations both public and commercial that are seeking for a better and faster means to forecast crime and a way to minimize the occurrences of crime are detailed in this research papers.

Keywords: Machine Learning, Crime Analysis, Pre- Processing, Crime Data, Arima Model, Kaggle, Data Mining,KNN(K-Nearest Neighbour) Data Security, Supervised Learning, Unsupervised Learning,Crime Response Technology, Smart Policing, Machine Learning, Natural Language Processing, Decision Trees, Linear Regression, K-Means Clustering.

I. INTRODUCTION

Crimes are waiting to happen in the shadows. But what if we could foresee them and take steps to stop them before the event occur? Data science has the ability to completely change how law enforcement agencies approach criminal investigations. Predictive modeling offers a method that can both develop a model and generate predictions, it was utilized to make predictions. This approach consists of many algorithms that can examine characteristics from the training data used to generate predictions.

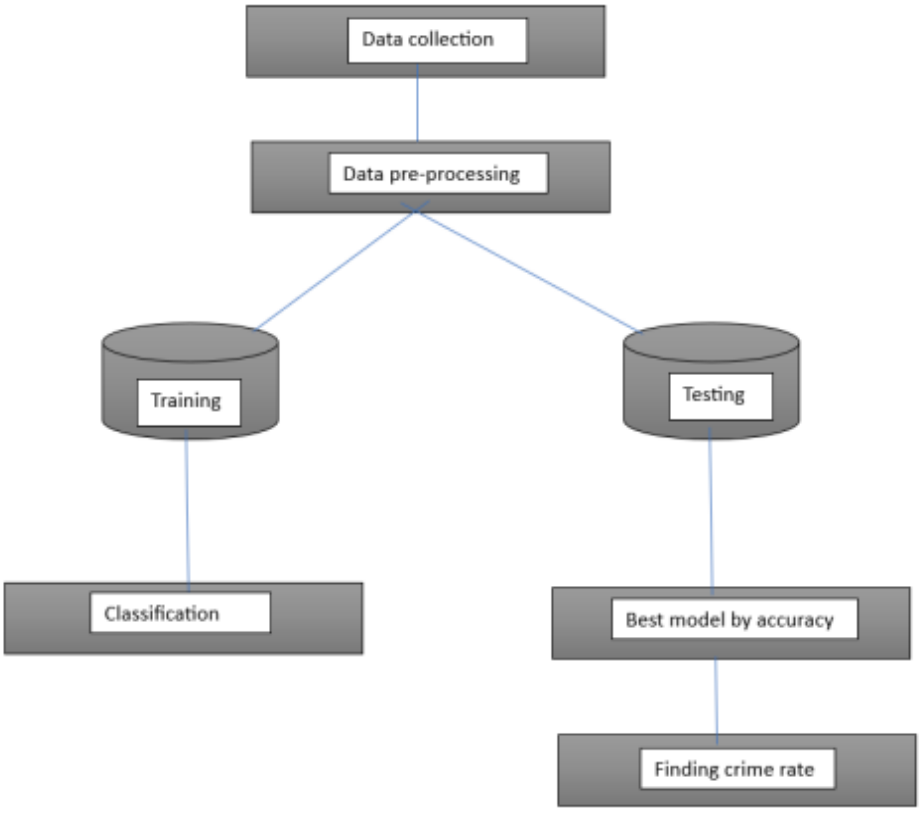

Fig. 1. Block diagram of crime data analysis

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Steps commonly in the papers selected:

1. Collect a crime data from any valid source
2. Data pre-processing: After collecting the data, we must delete any undesirable information [4] from the data set [5], such as null values, undesirable rows, and undesirable columns.
3. We must divide the data into training data and testing data after pre-processing it.
4. By training the data and testing the data against algorithms, we must test algorithm
5. Algorithms models to discover the best among them with the best accuracy.
6. Calculation of accuracy

The following are some advantages of this paper:

➔ Data preprocessing from the raw data makes the subsequent operations simple.

➔ We can gain a sense of the results of prediction using a graph or accuracy rate.

II. RELATED WORK

While predicting criminal activity, data scientists are like investigators, collecting all the evidence they can find and using it to solve cases. They collect crime data from all kinds of sources, clean it up and organize it so you can start predictive crime analysis. Find patterns and connections in your data using powerful algorithms and machine learning techniques. Once you have all the information you need, you will be able to predict where future crimes will occur.

A. Crime Analysis Mapping, Intrusion Detection – Using Data Mining[2020]

[6] An important aspect of crime analysis is data mining. The virtual identifier, the pruning method, the support vector machines, and the apriori algorithms are only a few of the many distinct algorithms that have been discussed in past research publications. VID seeks to connect the record and video. The fuzzy association rules algorithm and the apriori algorithm work together to identify a postal bomb attack in around 600 seconds. In this study, we found that this procedure might be sped up by using the KNN (K - Nearest Neighbour) and ANN (Artificial Neural Network) algorithms for [7] crime mapping analysis. The Office of Community Oriented Policing Services (COPS) manages and funds crime mapping. Research that is based on evidence aids in the analysis of crimes. Using data mining techniques, we determine the crime rate using the historical data. Crime analysis employs analytical methods in conjunction with both quantitative and qualitative data to solve crimes. The crime mapping is a crucial study subject to focus on for the sake of public safety. With the use of data mining algorithms, we can identify the areas where crimes occur the most frequently. We take the following actions in Crime Analysis Mapping to lower the crime rate:

1) Obtain crime statistics
2) Data in groups
3) Grouping
4) Data forecasting

Data mining technologies are used to reduce and prevent crimes and criminal disorders, and crime mapping aids in comprehending the theory and practice of crime analysis. [8] ANN (Artificial Neural Networks) and KDD (Knowledge Discovery in Databases)-based data mining techniques are available. Here they have gathered information from the police and make every effort to get specifics such as the person's name, height, age, sex, fingerprint information, and pattern identifying number for instances of a similar kind. Once we get the data, we begin the data processing. Along with the basic data, we receive a tonne of other information. To lessen or prevent confusion, we must first identify useless data and delete such types of data before analyzing the data using data mining techniques and tools. To find the patterns in the crime data, we employ the SAM tool. Here, the data are divided into two categories :

i. Supervised data ii. Unsupervised data.

By utilizing this supervised data to train, we take the data that contains all of the case's specifics and attempt to solve the other instances. We primarily gather data on traits, such as eye color, fingerprint details, characteristics, measurements, or other factors. The engineering, or model, the learning calculation, and the

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

enactment capacities are the three main components of a neural system. We make use of the KDD, which is an information-based learning revelation. This process involves the extraction of the exciting information that suggests the material should not be repeated and should now be clear. Information mining is the KDD's core component. The following are the methods used in the KDD: data cleansing, data storage, crime-related data storage, design, and assessment. With these methods, we will finally obtain the needed data.

B. Crime Data Analysis Using Machine Learning[2022]

We can find [5] the patterns [9] of criminal activities, by using machine learning. This study makes use of a crime data collection to forecast the sorts of crimes that will occur in a certain location, which speeds up the process of classifying criminal cases and moving forward as necessary. This essay makes use of facts from the last 18 years that were gathered from several reliable sources. The pre-processing of data is just as crucial as the final prediction; in this study, the data were cleaned and nourished using feature selection, null value removal, and label encoding. This study provides a useful machine learning model for forecasting the subsequent criminal case.

The study [10] 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. In this study, the models of linear regression, decision tree, and random forest are employed to predict crimes since they have a solid track record of accuracy. This study's dataset was obtained from data.world.com. According to the state and year, the data set includes various crime kinds that are perpetrated in India. The region where crimes are committed is the outcome of this article, which uses crime kinds as its input. Data cleansing, feature selection, and eliminating nulls are all part of the data pre-processing. Data cleaning, feature selection, removing null values, and data scaling via normalizing and standardizing are all part of the data pre-processing process. Following data preparation, the data is free of null values that might dramatically influence the model's accuracy, and feature selection is used to choose just the necessary characteristics that won't affect the model's accuracy. The chosen models, such as Logistic Regression, Decision Tree, and Random Forest, are trained by dividing the data into train and test data after data pre-processing. The great accuracy of the suggested system's model prediction approach is a benefit.

System implementation:

1. Upload dataset: Kaggle is where we find the sensor data.
2. Data pre-processing: after obtaining the data, it is necessary to eliminate any unnecessary information from the dataset, such as null values, unnecessary rows, and unnecessary columns.
3. Train and Test Model: Following pre-processing, the data must be divided into training and test sets.
4. Run Algorithm: Using data for training and testing, we must test machine learning algorithms to determine which one has the highest accuracy.
5. Accuracy graph: To finish, we show the accuracy graph for each method.

The process of testing involves running a programme with the goal of identifying errors. Our programme must be error- free in order to function properly. The software will be free of any errors if testing is successful.

Types of testing:

1. White Box Testing
2. Black Box Testing
3. Unit testing
4. Integration Testing
5. Alpha Testing
6. Beta Testing
7. Performance Testing

Basic information on criminal behavior in a neighborhood obviously contains markers that machine learning agents will use to categorize a criminal activity given a place and date. [12] The dataset's unbalanced categories cause problems [13] for the training agent, but it was prepared to fix the issue by over- and under-sampling the

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

data. This paper presents a crime data prediction using a Colab notebook with Python as its core language. Python provides built-in libraries like Pandas and NumPy through which the work will be completed more quickly, and Scikit provides all the processes of how to use different libraries providing by the python.

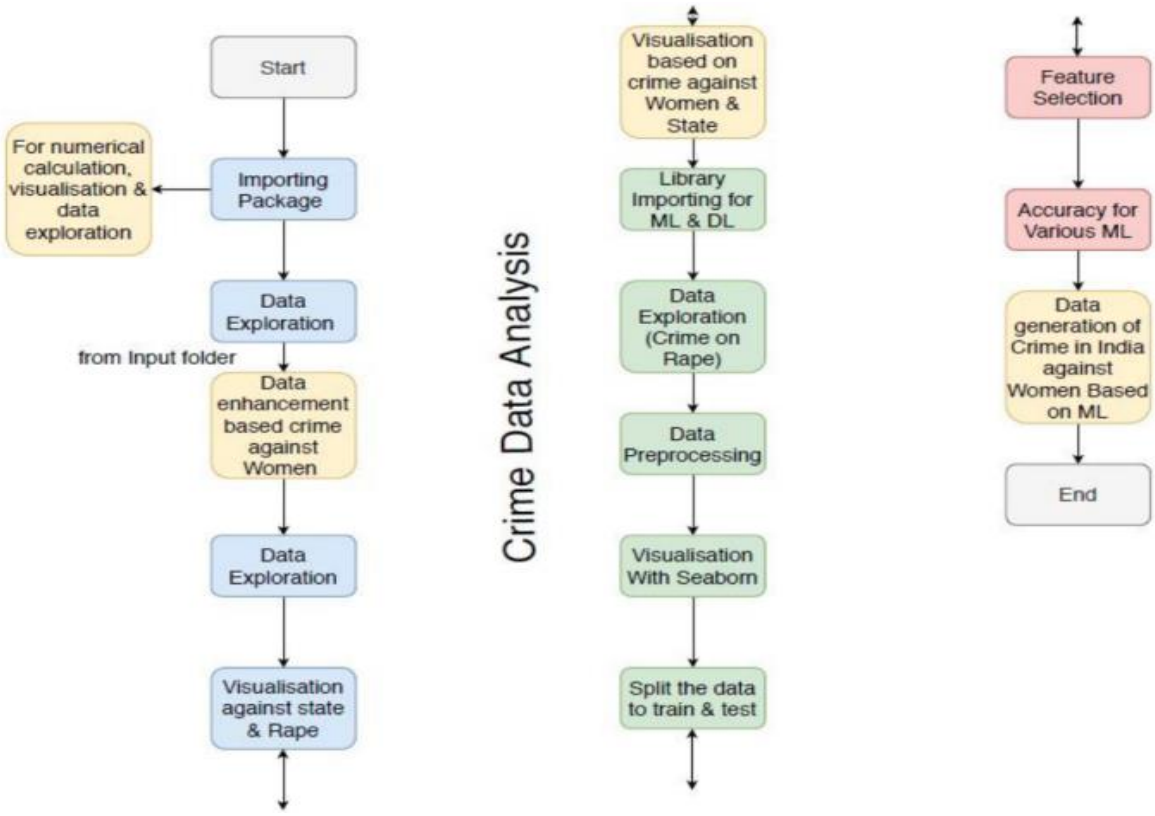

Fig. 2. Flowchart

C. Crime exploration and forecast[2022]

[14] This paper's goal is [15] to analyze a dataset of multiple crimes and forecast the kind of crime that could occur in the future depending on various factors. In this project, we will use machine learning and data science techniques to forecast crime using a set of crime data from India. A rigorous approach to identifying crime is through crime analysis and prediction. This system is able to foresee and highlight crime-ridden locations. We can draw previously undiscovered, important information from unstructured data using the concept of machine learning. We anticipate extracting additional data from the available databases. Crime is a dangerous and pervasive social problem that has an impact on individuals all around the world. Crime affects people's quality of life, prosperity, and safety. The following are the main objectives of crime evaluations: 1. Identifying crime patterns by studying actual crimes and criminal information. 2. Using geographic distribution to anticipate crime of available information as well as prediction of crime total using various data mining approaches. 3. Detecting crime A web-based application powers the system under evaluation. a sophisticated criminal detection system with the primary objective of predicting crimes and their trends. The suggested method uses automation for early crime pattern prediction, which helps deter crimes, as well as a data mining technique known as "Prediction Rules" for [16] crime pattern detection. based on information about previous crimes, the time and place, predicts crime patterns. 17] A regression method is developed to find the historical relationship between an independent and a dependent variable in order to estimate the dependent variable's future values.

Fig. 3. context diagram

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Regression makes predictions about future behavior of variables based on their past relationships. K-means clustering was used in the investigation of high- and low-frequency crime locations.

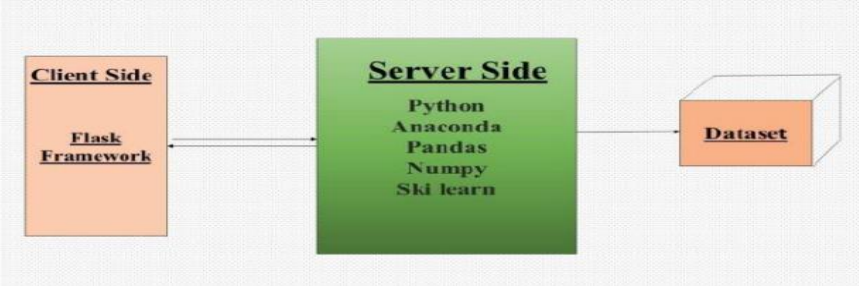

Fig. 4. Architecture design

D. Crime Rate Prediction Using Machine Learning and Data Mining [2021]

[18] In this work [19], we analyze Bangladesh's crime rate using various clustering methods of data mining, and we also train our data set using the K-nearest neighbor (KNN) algorithm. We use both primary and secondary data for our work. We can calculate the prediction rate of the path by using the algorithm to analyze the data and estimate the prediction rate of many crimes for many distinct locations. Finally, we utilize the projected rate to determine our safe route. This position will help people become aware of the crime hot spot and find a safe route to their destination.

Methodology used here:

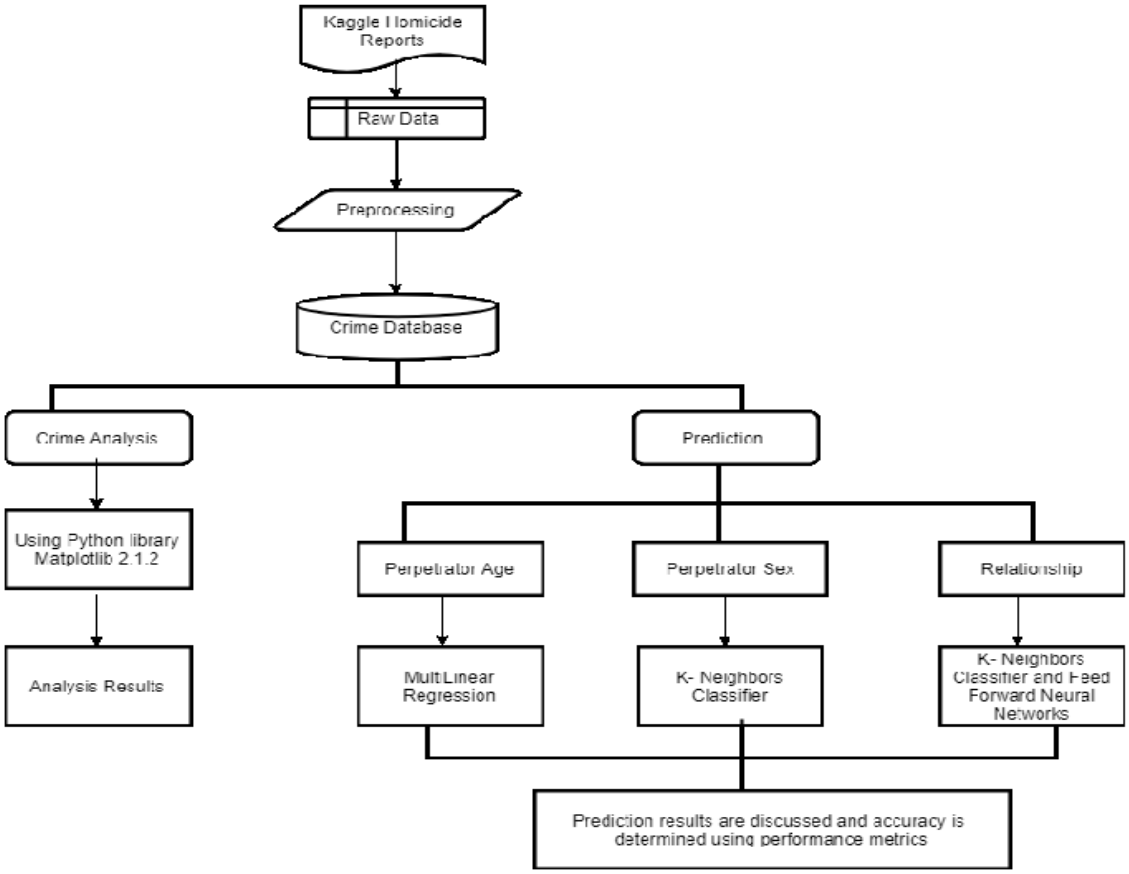

Fig. 5. Workflow diagram

There are four projected target variables.

i. To determine the accuracy of crime data based on age, linear regression is performed.

ii. The sex is calculated using the K-Nearest Neighbors classification.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

iii. The categorization of K-nearest neighbors is used to assess the perpetrator's gender.

iv. The KNN algorithm's final prediction rate for years depends on age, sex, time, and year.

There are several clustering methods in the subject. The K- means partitioning approach is widely used and accepted. We used the Linear regression technique instead of the K-means approach since it allows users to calculate the number of clusters based on those values Although Naive Bayes appears to produce decent results.

This research summarizes and informs the potential crimes based on crime data sets and algorithms. We determine the crime rate in many categories, such as age-based, male-to- female, area-based, and monthly crime rates. Various types of crime statistics, data from general population surveys, literature studies, and statistical models that project crime patterns into the future are some of the data sources and methodologies used to inform forecasting. By extending the results of a time series study of crime patterns into the future, algorithms that explain the behavior of observed past values may be used to predict future crime trends. Any predictive model aims to demonstrate the connection between a certain predictor and a dependent variable. To achieve higher accuracy, such models must recognize and forecast the range and make predictions about the kinds of variables that may affect victimization and crime in the future.

III. COMPARATIVE STUDY

In the first research paper, identification of Crime mapping analysis is based on KNN (K – Nearest Neighbor) and ANN (Artificial Neural Network) algorithms by classifying the data into supervised and unsupervised data. Law enforcement authorities can forecast future criminal activity after analyzing the data. These forecasts, which may take many different forms, aid organizations in more efficient resource allocation and crime prevention. The forecasts may take many different forms, including:

- Crime-ridden places
- criminal offenses that are most likely to occur
- when crimes are most likely to happen during the day
- repetition of some people's offenses
- Trends in crime patterns and criminal hotspots [20]

With the assistance of KNN , high probability of crime rate can be predicted and with the help of ANN we can analyze the observed data. Crime Mapping is conducted and Funded by the Office of Community Oriented Policing Services (COPS) [21]. Whereas in the second paper, Linear Regression, Decision Tree, and Random Forest models are employed in this study to predict crimes which have shown excellent accuracy. With the help of Data Mining technique , a crime predicting model was implemented along with a variety of interactive visualization tools and criminal networks to help in the investigation of crime in the third research paper. Here Crime report databases can produce important insights using data mining techniques. The initial stage of the study of crime is criminal analysis. Criminal analysis is the investigation of correlations between various crimes and criminal factors as well as their interactions. This study helps with the on-demand generation of data, queries, and maps. It also helps in figuring out whether a crime has happened in a certain, well-known place. Moving to the fourth paper, it used the algorithm of linear regression, Naive Bayes, and K-nearest neighbor. Among these K-nearest neighbors was chosen for crime prediction technique since it generally gives the appreciated accuracy. In this task, we analyze crime rate using several clustering techniques from data mining, and we also train our dataset using the K-nearest neighbor (KNN) method. In order to do our work, we use primary and backup information. We can calculate the prediction rate of the path by using the algorithm to analyze the data and estimate the prediction rate of many crimes for many distinct locations. Finally, we utilize the projected rate to determine our safe route. People will be helped by this employment to become aware of the crime area and find a safe route to their destination. Comparatively the second and fourth paper has excellent accuracy in their outcome.

IV. CONCLUSION

[22]The interesting subject of data science use maths and technology to reveal the mysteries of the data! Data scientists are like detectives when it comes to anticipating illicit activity; they gather all the available

information and use it to crack the case. They gather data on crimes from a variety of sources, clean it up, and arrange it before beginning their study for crime prediction. They look for patterns and links in the data using potent algorithms and machine learning approaches. It seems as though they have a special decoder that enables them to see what is happening in the background. They then generate forecasts about potential crime hotspots after they have all the necessary data. For law enforcement organizations, using data science to forecast criminal activity may be a game-changer. Imagine being able to prevent crimes from occurring in the first place! However, enormous authority also entails great responsibility. Particularly when it comes to delicate issues like privacy, it is critical to consider the advantages and disadvantages. Despite these difficulties, data science has promise for improving crime prediction. We can make communities safer and continue to be at the forefront of the battle against crime by using it responsibly.

We employ K-nearest neighbor for our crime prediction technique since it generally gives the appreciated accuracy despite some linear operating well and offering superior precision. In the future, we will be able to forecast events with more accuracy to these technologies, and we will also be able to detect and locate criminal hotspots thanks to this precision. Law enforcement authorities are able to forecast future criminal activity with the use of crime analysis. These forecasts, which may take many different forms, aid organizations in more efficient resource allocation and crime prevention.

ACKNOWLEDGEMENT

We extend our sincere gratitude to all the teaching staff, Department of Computer Science, Vimala College, Autonomous Thrissur, for their valuable guidance and support at each stage of the project. We are also thankful to our parents for the support given in connection with the review.

V. REFERENCES

[1] B. Stickle and M. Felson, “Crime rates in a pandemic: The largest criminological experiment in history,” American Journal of Criminal Justice, vol. 45, no. 4, pp.525–536, 2020.
[2] P. Ekblom, “Getting the best out of crime analysis,” 01 1988.
[3] C. C. K. Jenga and G. Ka, “Machine learning in crime prediction.” https://www.ijrpr.com/, Feb 2023.
[4] C. P. S. M, “Road accident prediction and classification using machine learning”, IEEE Xplore, 2022.
[5] RJMETS, “International research journal of modernization in engineering technology and science.” https://www.irjmets.com
[6] B. Panja, P. Meharia, and K. Mannem, “Crime analysis mapping, intrusion detection-using datamining,” in 2020 IEEE Technology & Engineering Management Conference (TEMSCON), pp.1–5, IEEE, 2020
[7] P. Kedia, “Crime mapping and analysis using gis,” International Institute of Information Technology, vol. 1, no. 1, pp. 1–15, 2016
[8] S. Borah, R. Pradhan, N. Dey, and P. Gupta, “Soft computing techniques and applications,”
[9] P. Bhanumathi and S. Greeshma, “Crime data analysis using machine learning,”
[10] Researchgate, “Research gate.” https://www.https://www.researchgate.net.
[11] GRIET, “Griet.” https://www.cse.griet.ac.in/.
[12] R. J. D. D. S. R. G. Soma Sekhar, Puvvada Abhinaya, “Criminality data scrutiny using logistic regression algorithm,” IEEE Xplore, 2023.
[13] Student paper of Bournemouth University, “Bournemouth university.” https://www.bournemouth.ac.uk.
[14] T. U. K. Roopa, “Crime exploration and forecast,” 2022
[15] Educoglobal, “Educo.” https://www.https://www.educoglobal.com.
[16] M. A. Khan, A. Nasir, M. N. Ali, U. Farooq, and S. A. Malik, “Crime detection using digital forensic technology,” International Journal of Computer Science and Information Security (IJCSIS), vol. 14, no. 10, pp. 487–506, 2016
[17] IJPRSE, “The international journal of progressive research in science and engineering.” https://www.journals.grdpublications.com.
[18] springer, “Springer.” https://www.https://https://link.springer.com/

[19] S. Mahmud, M. Nuha, and A. Sattar, "Crime rate prediction using machine learning and data mining," in Soft Computing Techniques and Applications: Proceeding of the International Conference on Computing and Communication (IC3 2020), pp. 59–69, Springer, 2021.

[20] S. Vijayarani, E. Suganya, and C. Navya, "Crime analysis and prediction using enhanced arima model," Journal Homepage, 2021.

[21] slideshare, "ptslideshare." https://www.pt.slideshare.net/, Dec 2013.

[22] R. M. Saeed and H. A. Abdulmohsin, "A study on predicting crime rates through machine learning and data mining using text," Journal of Intelligent Systems, vol. 32, no. 1, p. 20220223, 2023.