

An Adaptive Technique for Crime Rate Prediction using Machine Learning Algorithms

Dr. V. Chandra Shekhar Rao¹, Kallepelly Spandhana², Dr. C. Srinivas³, Dr. M. Sujatha⁴, Bojja Vani⁵, Dr. S. Venkatramulu⁶

¹Associate professor , Dept. of Computer Science and Engineering, Kakatiya Institute of Technology and Science Warangal, India.

¹vcsrao.cse@kitsw.ac.in

²M.Tech Scholar, Dept. of Computer Science and Engineering, Kakatiya Institute of Technology and Science Warangal, India.

²kallaepellyspandhana@gmail.com

³Associate professor , Dept. of Computer Science and Engineering, Kakatiya Institute of Technology and Science Warangal, India.

³cs.cse@kitsw.ac.in

⁴Assistant professor , Dept .of Computer Science and Engineering, Kakatiya Institute of Technology and Science Warangal, India.

⁴Drms.cse@kitsw.ac.in

⁵Assistant professor , Dept. of Computer Science and Engineering, Kakatiya Institute of Technology and Science Warangal, India.

⁵Vanitanish2018@gmail.com

⁶Associate professor , Dept of Computer Science and Engineering, Kakatiya Institute of Technology and Science Warangal, India.

⁶svr.cse@kitsw.ac.in

Abstract— Any country must give the investigation and preventive of crime top priority. There are a rising amount of cases that are still pending due to the rapid increase in criminal cases in India and elsewhere. It is proving difficult to classify and address the rising number of criminal cases. Understanding a place's trends in criminal activity is essential to preventing it from occurring. Crime-solving organisations will be more effective if they have a clear awareness of the patterns of criminal behavior that are present in a particular area. Women's safety and protection are of highest importance despite the serious and persistent problem of crime against them. This study offers predictions about the kinds of crimes that might occur in a particular location using ensemble methods. This facilitates the categorization of criminal proceedings and subsequent action in a timely manner. We are applying machine learning methods like KNN, Linear regression, SVM, Lasso, Decision tree and Random forest in order to assess the highest accuracy.

Keywords-crime rate; criminal cases; style; ensemble methods; machine learning.

I. INTRODUCTION

As one of our general societal most pervasive problems, crime prevention is a vital endeavour. Major issues that are alarming. The methodical investigation and identification of multiple linkages, patterns, and trends in crime and disorder is known as "crime analysis and prediction." Over the world, crime is a widespread occurrence, especially in urban areas. As a result, the social order of today's society is upended, and fundamental human rights are infringed. The sites with a high crime rate at a specific period are not always easy for law enforcement officials to identify manually. Due to the evolving nature of criminal impulses and the erroneous use of the existing crime data. With an increase in population, the patterns of criminal activity evolve and become more complex.

New approaches are required to understand it and deal with it. Many have experienced job insecurity and economic losses as a result of the current COVID period's statewide shutdown, which has resulted in a considerable deal of psychological trauma. As a result, burglaries, shoplifting, and several other criminal actions have escalated. Some criminals have gotten away with breaking the law, are now free to roam through society, and constitute a danger to it. Governments in the majority of modern countries maintain databases on these criminals that can be used to find them.

To determine a case based on specific evidence, a thorough independent inquiry and evaluation should be carried out. It is very difficult for the authorities to assess and make decisions on these criminal matters given the number of crime data that is currently available in India. The objective

of this research is to create a solution for the problem of crime decision-making after recognising it as a significant issue. The automobile starts to move independently. An autonomous driving vehicle repeatedly goes through the identification, judgement, and control stages as it makes different decisions to reach its goal. Identifying the criminal trend. Crime should be organised by place. Analysis of crime.

II. LITERATURE SURVEY

The objective is to define how the characteristics should be applied to segment the whole dataset; once this is done, various datasets will invariably fall as one or more groups. With the use of a unique information mining process called classification, every item in a collection of data is sorted into one of the specified classes or groupings. By categorising data, it is possible to make it simple to understand existing datasets and to forecast how additional new individual datasets will behave.

Data mining creates classification models by looking at categorised data and figuring out how those data are correlated. The Bayesian principle is used to make predictions using the naive Bayes categorization method. Rayhan et al. [1] did a study using spatial and temporal attention to predict the top 4 categories of crime. Their model is unable to forecast categories with little data. As crime prediction can help authorities.

The primary goal of time series analysis is the extraction of useful statistics and other properties from time-series data. Based on previously observed data, this information strongly predominates in predicting future values. Violent crime and mishaps are increasing as a result of population growth and ongoing urbanisation. One of the current fads involves using Big Data Analysis (BDA) to extract information from these enormous amounts of data and analyse hidden patterns and the co-relation between them [2].

S. Sathayadevaan, M. S. Devan, and Surya Gangadhaaran, introduces analysis and crime forecasting using data mining techniques and algorithms. The system categories the causes of crimes, such as criminal family information, additional motives, and crime records from the regular detect examination. It uses classification to determine the crime ratio. Model evaluation using the Naive Bayes method makes it possible to create a training programme for a variety of offences [3].

It is a tremendous undertaking to expose criminals and the vulnerable situations that are continually influenced by their illegal behavior in the modern digital age. One of the key tasks of law enforcement agents is the challenging analysis of a massive volume of data regarding crimes and

criminals. We have a variety of useful and practical tools for analyzing huge and diverse quantities of data thanks to data mining. To investigate, manage, and prevent crime, it aids organizations and people in locating concealed facts from a sizable database of criminal records.

A data mining approach that can aid in more quickly solving crimes can be developed by combining computer science with criminal justice. This paper's main objective is to describe different categorization techniques. Feed-forward the most effective time-series forecasting technique is an ANN or multilayer decision tree because it doesn't need to know how the data will be distributed beforehand [5].

Sunil Yadav et al. used the apriori algorithm, K-means, Naive Bayes, and Linear regression approaches to improve the accuracy of crime prediction. The study's data sources came from the police departments of several Indian states. [6]

Instead of concentrating on causes of crime such as criminal histories of offenders, political enmity, and so on, Sathyadevan et al. have focused mostly on everyday crime elements. Data is gathered via a number of different sources, including social media platforms, websites, media websites, and RSS feeds. By training the naive Bayes classification algorithm on crime data, including burglary, murder, and vandalism among others, a model is created. More than 90% of crimes were correctly predicted using the Naive Bayes algorithm.

III. METHODOLOGY

The project is completely based on the crime rate prediction using machine learning and ensemble techniques

A. Clean data set

Data cleaning is the process of removing faulty, redundant, or other incorrect data from a dataset. When two or even more databases are combined, these errors typically arise and can consist of incorrectly labelled data, redundant entries, improperly formatted data, and other issues. With data cleansing, the quality of your data is increased

B. Separating the data into test and training sets

By dividing the values into two parts and training a subset of it, it is possible to create a trained (fitted) line that can be used to estimate new or unknown datasets with accuracy. This method will help generalise how unfamiliar and new data (the test set or new dataset) will be assessed. The train-test divide ratio is often 70:30 or 80:20

Scaling down the learned model: It is a technique used to normalize a variety of numerical variables with different

magnitude ranges. When features such as height, BMI, or age have varying magnitudes and units, or when some features have values in the tens of thousands while other features have values in the tens or hundreds of thousands, the value of each item to the dependant variable will vary.

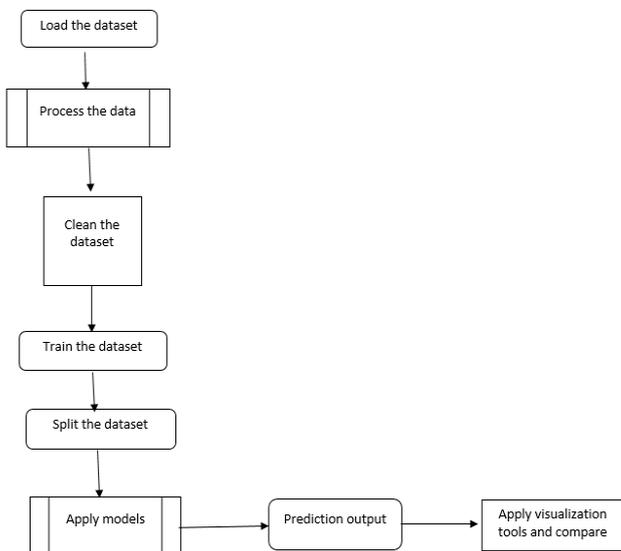


Figure 1. Flowchart for crime rate

C. Ensemble techniques

A well-liked strategy in machine learning for enhancing the predictive performance of models is assembling approaches. Ensembling techniques are a type of machine learning approach where predictions from various models are combined to provide a more reliable and accurate final prediction. Ensembling is based on the theory that integrating forecasts from various models can lower the danger of over fitting and increase the prediction's overall accuracy. Ensembling is a technique that can be used to solve a range of machine learning issues, such as classifying, regression, and clustering. Below are several ensemble methods that can be applied to crime rate prediction:

❖ Linear Regression

It is a method of conducting predictive analysis using statistics. Linear regression generates predictions for variables that are actual, continuous, or numerical. Importing the necessary libraries, such as sea born and matplotlib for data visualisation and pandas and numpy for data analysis and manipulation. The linear regression approach would look for the best-fit line that best describes the connection between the distinct variables and crime in the case of predicting crime rates. Based on these features, the line of perfect fit could be utilized to forecast rates of crime for new sites.

It is significant to remember that regression model is a linear equation and assumes a linear connection between both the variables that are dependent as well as independent. Data must

be cleaned and altered to meet the requirements for exploratory data analysis to be performed, including handling any null values, updating to the appropriate formats, changing data types where necessary, eliminating unneeded rows or columns, etc. Before evaluating the raw data for visualisation, it must be thoroughly cleaned of any muck, regardless of how you obtain it.

❖ Lasso Regression

An approach to regularisation is lasso regression. For a more precise prediction, it is utilised instead of regression techniques. It makes use of shrinking. Data values shrink towards a mean, which is a centre point, in shrinkage. The lasso approach promotes straightforward, sparse models (i.e. models with fewer parameters). When a model exhibits significant levels of multi collinearity or when certain steps in the model selection process, such as variable selection and parameter removal, need to be automated, this specific sort of regression is a good choice.

L1 regularisation technique is used in Lasso Regression. Because it automatically performs feature selection, it is employed when there are more features. Lasso modeling can be used to narrow down a huge pool of potential variables in the context of predicting crime rates and find the most pertinent predictors. As an illustration, socioeconomic variables like the rate of poverty, the unemployment rate, and the degree of education could be thought of as predictor variables of crime rates. Some of these factors may have a strong correlation with one another, and not all of them may be important. Lasso regression can assist in identifying the most crucial factors that are highly related to crime rates and removing the others.

❖ SVM

Support Vector Machine (SVM) is typically used in machine learning to address categorization problems. Support Vector Machines are a powerful machine learning technique for forecasting crime rates (SVM). SVM, a technique for supervised learning, can be used for both regression and classification tasks. Once the dataset has been gathered, SVM can be used to train a model which can forecast crime rates depending on these variables. In SVM, the algorithm looks for the hyperplane that divides the data into distinct classes or accurately predicts the target variable.

The SVM model would look for the hyperplane that best divides the information into both high and low crime rate areas depending on the various criteria in the example of predicting crime rates. Based on these features, the hyperplane is going to be utilized to forecast rates of crime for new sites. Based on these features, the hyperplane is going to be utilized to forecast rates of crime for new sites.

❖ Decision Tree

Decision Trees are most typically employed to solve classification problems. Decision tree algorithm can be used to forecast crime rates by examining the numerous elements that influence criminal activity in a certain area or region. Demographic information like ages, gender, educational status, employment history, and income may be among these criteria, as well as environmental information like the number of police officers nearby, the severity of the local poverty, and the accessibility of social services.

Once trained, a decision tree can be used to predict future rates of crime according to modifications in the variables that were determined to be most crucial. For instance, if a decision tree determined that poverty was the primary cause of crime in a specific location, policy-makers may utilise that information to focus on anti-poverty initiatives.

❖ Random Forest Algorithm

The network entities that make it up a tree structure are a transformation to convert a set and a child nodes. Whereas Leaf nodes represent the results of choices and lack any more branches, Decision node are utilized to make these decisions and also have numerous branches. One advantage of using the random forest approach to predict crime rates is that it can take a large quantity of input data without overloading the model. Random forests are also rather impervious to extremes and missing information, which are frequent problems when dealing with crime data. While a random forests algorithm can be helpful for predicting crime rates, it should be combined with other techniques to gain a more complete overview of the variables influencing crime in a certain location.

The Random Forest Algorithm's capacity to handle data sets containing both dependent variable, as in regression, and categorical data, as in classification, is one of its most significant features. For jobs involving classification and regression, it performs better. In this tutorial, we'll learn how random forests function and use them to solve a classification problem.

❖ KNN

K-Nearest Neighbour is one of the most basic supervised learning-based machine learning approaches. On the basis that the present specific instance and the current instances are comparable, The K-NN algorithm assigns the new information to the group that most closely fits the pre-existing categories.

The K-NN method keeps the entire previously gathered data while classifying new data based on similarity. This means that utilising the K-NN technique, new data can be consistently and swiftly categorised into the proper category. Although

regression and classification problems can both be solved using the K-NN technique, classification problems are the ones where it is most often used.

IV. RESULT ANALYSIS AND DISCUSSION

Assembling processes refers to the act of combining the predictions of several machine learning models in the interests of enhancing the final prediction's precision. Ensembling approaches can be used to integrate the outcomes of various algorithms that have been trained on various datasets or using various algorithms in the context of rate of crime predictions to get a more precise and reliable prediction.

As that crime is a complex issue driven by a variety of elements, including demographic information, economic data, and environmental conditions, assembling methods can be very helpful in predicting crime rates. Ensembling techniques can assist in producing a more precise and thorough prediction of crime rates by merging the predictions of many models trained on various subsets of these elements. Using data pre-processing and categorization procedures, it is difficult and requires extensive understanding of numerous statistical indicators to analyze and interpret communities and crime data sets. The amount of training time was lowered by using the transfer learning technique.

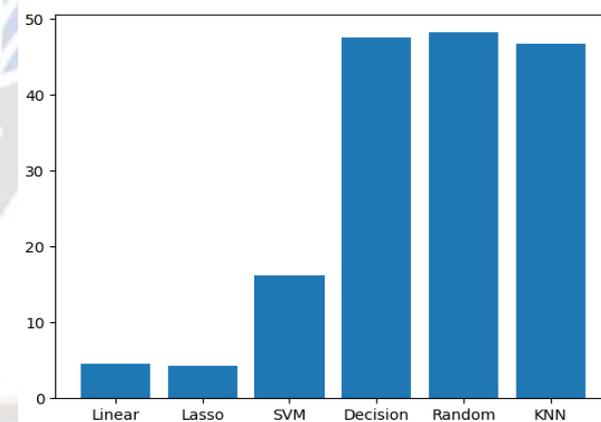


Figure 2. Comparison between different methods

It is beneficial for the authorities and law administration organisations to design new legislation for certain offenses that are probable in multiple states and regional cities because of algorithm forecasts distinct crimes per each district and state. Several elements influence criminal behaviour. For instance, when people fear they will not have sufficient resources to survive, they could resort to severe measures, which increases crime rates in areas of poverty. Together with revenge, the thrill of committing. In addition to criminality, mental illness, or an obsession, there are additional causes of criminal behaviour that may exist. The predictions made by various classification algorithms show the possibility that a crime will be committed,

including if it will be violent or not, whether a murderer would conduct it or not, and even if it will be committed at all. These predictions could help the FBI and local authorities solve a number of cases quickly and accurately.

The Random Forest Algorithm, which creates a classifier model for every feature and makes decisions by a majority vote, is the most successful classification algorithm. In an ensemble learning method known as a Random Forest, a large number of decision trees are constructed during training to create a class that represents the average of their classification or regression results in terms of performance and responsiveness. By training various decision trees on a wide range of data points and combining their predictions, this approach can be used to estimate the crime rate.

V. CONCLUSION

The implementation of the prediction-rate area-specific modeling is complicated by the absence of crime in many areas. In that work, we created and tested a crime prediction model using age, gender, season, moment, and month data. The predictions given by several classification algorithms demonstrate the likelihood that a crime will occur, including whether it will be violent or not, whether it will be committed by a murderer or not, and if it will be committed at all. These forecasts could aid the FBI and local police in efficiently and accurately resolving a number of cases. The most effective classification algorithm is the Random Forest Classifier, which builds a classification tree for each feature and decides using a majority vote. High accuracy can be attained via the Random Forest Classifier. Also, we observed that the dataset performs well with non-linear data as compared to linear data hence not so good results were achieved with Linear Regression.

REFERENCES

- [1] Y. Rayhan and T. Hashem, "AIST: An interpretable attention-based deep learning model for crime prediction", arXiv: 2012.08713, 2020.
- [2] M. Feng, J. Zheng, J. Ren, A. Hussain, X. Li, Y. Xi, et al., "Big data analytics and mining for effective visualization and trends forecasting of crime data", *IEEE Access*, vol. 7, pp. 106111-106123, 2019.
- [3] S. Sathyadevan, M. S. Devan, and S. Surya Gangadharan, "Crime analysis and prediction using data mining," 1st Int. Conf. Networks Soft Comput. ICNSC 2014 - Proc., pp. 406–412, 2014, doi: 10.1109/CNSC.2014.6906719.
- [4] Farhad Khoshbakht, Atena Shiranzaei, S. M. K. Quadri. (2023). Design & Develop: Data Warehouse & Data Mart for Business Organization. *International Journal of Intelligent Systems and Applications in Engineering*, 11(3s), 260–265. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/2682>.
- [5] M. C. Bishop, "Multi-layer perceptron," in *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995, pp. 116–163.S.
- [6] Yadav, M. Timbadia, A. Yadav, R. Vishwakarma and N. Yadav, "Crime pattern detection, analysis & prediction," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, pp. 225-230, doi: 10.1109/ICECA.2017.8203676.
- [7] Sukumar P, Robert L, "A Probe on Crime Data in Various Domains," *International Journal of Recent Technology and Engineering*, ISSN: 2277-3878, vol. 8(6), pp. 3940-3948, Mar. 2020, doi: 10.35940/ijrte.F8951.038620.
- [8] Ms. Sweta Minj. (2012). Design and Analysis of Class-E Power Amplifier for Wired & Wireless Systems. *International Journal of New Practices in Management and Engineering*, 1(04), 07 - 13. Retrieved from <http://ijnpme.org/index.php/IJNPME/article/view/9>.
- [9] 2020 IEEE Technology & Engineering Management Conference (TEMSCON) "Crime Analysis Mapping, Intrusion Detection Mining".
- [10] R.M. Aziz, p. Sharma and A. Hussain, "Machine learning based soft computing regression analysis approach for crime data prediction," *Karbala International Journal of Modern Science*, vol.8, no. 1, pp. 1-19, 2022.
- [11] Joseph Miller, Peter Thomas, Maria Hernandez, Juan González, Carlos Rodríguez. Exploring Deep Learning Models for Decision Science Applications. *Kuwait Journal of Machine Learning*, 2(3). Retrieved from <http://kuwaitjournals.com/index.php/kjml/article/view/198>.
- [12] S. Ramasubbareddy, T. A. S. Srinivas, K.Govinda and S. S. Manivannan, "Crime prediction system," *Innovations in Computer Science and Engineering*, vol. 103, no.3, pp. 127-134, 2020.
- [13] S. Rishabh, R. Rishabh, K. Vidhi and C. Prathamesh, "K-means clustering analysis of crimes on Indian Women," *Journal of Cyber Security and Information Management (JCIM)*, vol. 4, no. 1, pp. 5-25, 2020.