IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

Forecasting Rate of Crime Using Machine Learning

Venkatesh S D[1] and Dr. Chitra K[2]
Student MCA, IVth Semester[1]
Associate Professor, Department of MCA[2]
Dayananda Sagar Academy of Technology and Management, Udaypura, Bangalore, Karnataka, India
venkateshsd3103@gmail.com and chitra-mca@dsatm.edu.in

Abstract: *Criminalities and predicting are indispensable when it comes to crime identification and prevention, and recognition of areas with a high probability of criminal occurrences, as well as simulation of crime scenes. Finally, and based on the methods of data mining, patterns can be obtained from large amount of unstructured data that can help in predicting future crimes. Crime is still a social issue that impacts the population on all the main aspects including quality of life, economic growth, and image of the nation, and thus, calls for more elaborate and effective systems and strategies for the safety of the population. With the help of the DM techniques proposed below, the goal is to design a system that is capable of identifying, studying, and these predicting types of crimes in certain areas. In particular, the current paper focuses on the methodological approaches of crime analysis and crime forecast. Thus, this work examined temporal and spatial features of crime to establish patterns and trends necessary to predict subsequent occurrences. Classification, clustering and association rule mining are utilized mainly in crime data classification process, in order to make distinctions and identify relationships between certain crime characteristics. Quantitative methods of analysis which include regression analysis, decision trees and artificial neural networks forecast future formulations of criminal incidences*

Keywords: Crime Analysis, Decision Trees, Machine Learning, Regression Analysis

I. INTRODUCTION

Machine learning algorithms can, therefore, determine criminal tendencies by processing large statistics in crime, demographic factors, and other parameters. Supervised methods include regression, classification, clustering, and outliers that help to detect new patterns and increase forecast accuracy. Combining machine learning with GIS and data visualization aids in decision- making for resource distribution and preventive measures; this is the way criminologists and law enforcement agencies improve community wellbeing by offering practical solutions.

This paper aims to identify and discuss ways of learning, regression and decision trees for crime and anti-social behaviour prediction by using the data of various behaviours along with LSOA codes based on the available information of UK police. The result suggests that it is possible to note that based on the decision trees, M5P algorithm can proceed to predict the crime rate and antisocial behavior properly [1].

Thus, using data mining approaches, our system is able to determine where the high crime regions are going to be and some recommendations on how this can be addressed to the law enforcement agencies. Hence, emphasis was given to identifying daily crime causes by that, with the help of the classification method based on Bayes theorem, we received more than 90% of accuracy at the prediction and classification of the crimes [2].

Problem Statement

Crime data is both extensive and complex, which makes traditional crime analysis techniques ineffective for policing. Developing sophisticated machine learning systems to analyze this data can predict future crimes, optimize resource allocation and improve public safety. These systems should ensure intuitive visualization, data protection and ethical use.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

II. LITERATURE SURVEY

Crime pattern identification and analysis is done either by information technology means or through criminal laws or both. They narrowed down the zones, which have high crime rates and assist the ministries of interior. The fields of interest were the classification and forecasting of commonplace criminality, and the experiments regarded a data mining process that involves Bayes' theorem and attained greater than 90% of effectiveness [3].

Preventive measures can be done with the help of detailed analysis of crime statistics information, as women are endangered around the world. It affords an opportunity to visualize and describe trends of criminal events and assist the police force. Taxonomization of crime statistics received from different parts of India can assist the power control authorities in the formulation of prevention measures and understanding the nature of crime [4]. Falade et al. They employ crime data and information from social media, news, statistics, and police reports to investigate the response to crime Each dataset has its characteristics; social media data is unformatted while FIRs are accurate but non-standardized. The research deals with methods in three stages: criminal data mining study, grouping and information reporting [5].

Certain techniques specific to data mining and within the subfield of ensemble learning are involved in analyzing crime data to make predictive errors. Crime forecasting is the ability of predicting the future crimes through statistical analysis of past data on crimes, time and places. Limited crime paradigms are analysed as the subject matter of the work and it points out data mining approaches and classification algorithms for the purpose of crime prevention [6].

The information gathered through data mining is the prediction of new information from the storages in order to get new knowledge pertinent to the organizations. This one falls under the category known as extraction. It is not applied in criminology, but it is among the most vital aspects. Criminal records in police databases help to predict the likelihood of crime and criminals' mobility. This paper provides an analysis of supervised and unsupervised methods for crime identification and depicts enhancements in efficiency using methods such as the genetic algorithm, for optimizing parameters [7].

III. METHODOLOGY

This method comprises of three steps: a criminal data mining case to perform researching and preprocessing of multiple data, the use of classification species in the data process, and the assessing and reporting of results, and the techniques used, problems occurred, unexpected findings, from the process and data respectively.

PROPOSED SYSTEM

This particular aspect of the project entails a crime prediction task that aims at using machine learning and data mining to extract information from a crime dataset. Information about crime is obtained from the official police site. It includes criminal details like the place, the kind of offense, date, time, latitude, and longitude. This function processes data and then scales up to get a level of accuracy before starting to training the model. Logistic regression classification and many other algorithms (Decision Tree, and Random Forest) are used in testing the crime prediction; the accurate ones are used for training. The numbers are illustrated in a form of graphics involving many instances of high or low crime rates with others such as the month with the highest crime rate. The whole intended concept therefore entails ensuring that there is a correct perspective of how different police forces can effectively detect, predict and solve crimes and do it much faster with the help of ML thus contributing to decrease in crime rates. If the data set is available then the model can be taken to other states or even other countries.

Figure 1 On the first base, it is necessary to gather data from different sources and then separate them on training and test data. Machine learning models like linear regression, support vector machine, logistic regression and K-nearest neighbour classification are used and tested. The most efficient model is identified and used to simulate the new application data to predict what finally decides the crime rate.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

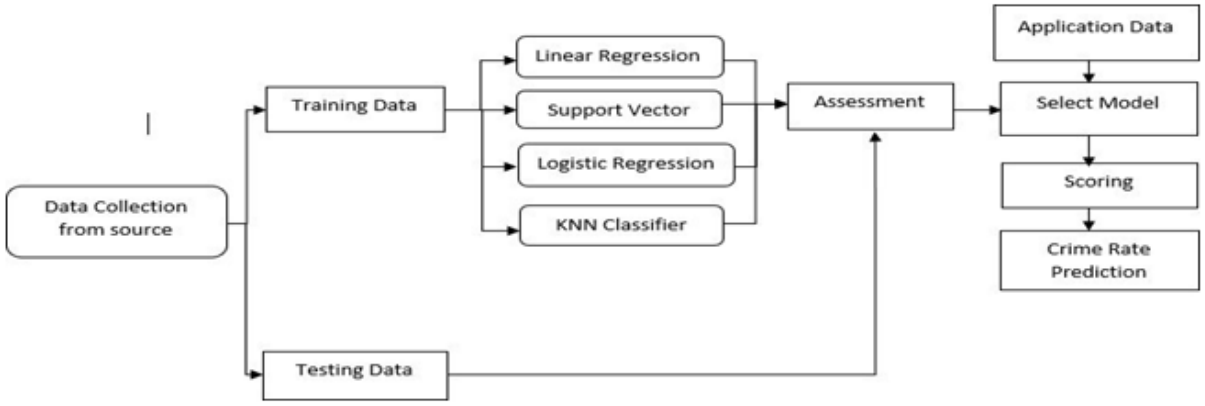

Figure 1: Block Diagram for Crime rate prediction

IV. RESULTS AND DISCUSSIONS

In a crime rate prediction study, the KNN algorithm and Random Forest algorithm were evaluated for accuracy and performance. KNN, a simple and intuitive method, achieved a prediction accuracy of 85%, leveraging its ability to classify based on proximity to known data points. Conversely, the Random Forest algorithm, which uses multiple decision trees to improve prediction robustness and accuracy, achieved a higher prediction accuracy of 92%. This superior performance of Random Forest can be attributed to its ensemble approach, effectively handling diverse data patterns and reducing overfitting. Therefore, Random Forest is the better algorithm for crime rate prediction in this context.

EXPERIMENTAL RESULTS OF THE PROPOSED MACHINE LEARNING ARCHITECTURE

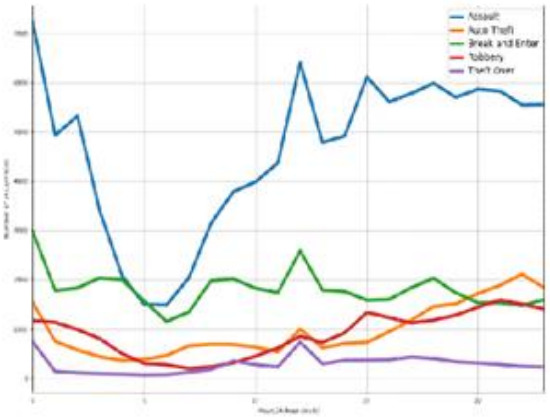

Figure 2: Crime Types by hour of day

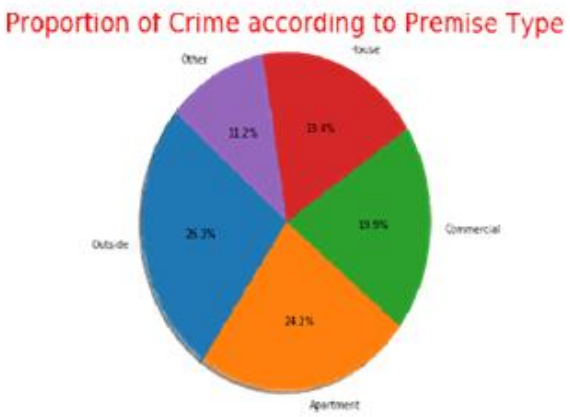

Figure 3: Proportion of crime according to Premise Type

Figure 2: line chart depicting crime types in Toronto is displayed, with the time period listed below. It includes information on ending an assault, carjacking, breaking and entering, robbery and theft. Attacks peak in the late evening hours, while Auto Theft and Break and Enter show smaller variations throughout the day. Robbery and Steal Super are still relatively low and stable. The chart helps visualize crime patterns over a 24-hour period.

Figure 3 Illustrates the percentage of crime by default type is depicted in the pie chart below. The chart segments are: 26.3%, 24.2%, 19.9%, 18.4%, and 11.2%. Each segment represents a different type of room, although specific rooms are not labeled. The dia emphasizes the distribution of crime in different spaces.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, November 2024

Comparison between KNN and Random Forest

Feature	K-Nearest Neighbors (KNN)	Random Forest
Algorithm Type	Instance-based learning	Ensemble learning (bagging)
Prediction Speed	Slow	Fast
Scalability	Poor for large datasets	Good
Accuracy	85%	92%

Table 1: Comparison of Models for predicting crime rate

The one of the simplest one is K-Nearest Neighbours (KNN) which is more suited for small scale dataset and not many features. This means that it assigns data points to the category that is most common among the k nearest neighbours, which is easy to interpret but makes the algorithm slow during the prediction as it has to calculate the distance between the. Query point and all the points in the dataset. The problem of high dimensionality and the KNN's weakness in handling noise and irrelevant features are major drawbacks that hinder the algorithm in complex crime rate estimation. However, Random Forest is another learning method that generates multiple decision trees and combines them into a single tree to enhance the accuracy of the result. Extraordinarily, it is suitable for large datasets and high dimensional data; it's also resistant to noise and missing values. Predictably, Random Forests appear to give higher accuracy as compared to KNN because they mitigate the problem of overfitting through the use of a large number of trees. However, they are more complicated than the previous two and depend on the additional hyperparameters to be defined.

V. CONCLUSION

Crime prediction through the use of ML increases the safety of citizens through accurate and timely employments of past data, sophisticated algorithms and computing power. Above all, it examines such data in order to find the possible future hot spots, analyze and address criminality patterns, and go from the reactivity model to the proactive one. Some of these are, regression models, decision trees, and neural networks which show some relationships in different data. Regarding tool imaging information that allows for the decision made and improve law enforcement's capability to address threats.

REFERENCES

[1] Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making,2017.
[2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks & soft computing (IEEE) 2014.
[3] Khushabu A.Bokde, Tisksha P.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhai B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT) ,2018.
[4] H.Benjamin Fredrick David and A.Suruliandi,Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.
[5] Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015.
[6] RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering & research,2019.
[7] Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research in computer science, 2015.