



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 12, December 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Criminal Data Analysis Using Machine Learning Models

M Durga Bhavani¹, Firoze Pattan²

PG Scholar, Dept. of CSE, ST Marys Group of Institutions Guntur, AP, India¹

Assistant Professor, Dept. of CSE, ST Marys Group of Institutions Guntur, AP, India²

ABSTRACT: The criminal cases in India are increasing rapidly due to which number of cases pending are also piling up. This continuous increase in the criminal cases is proving to be difficult to be classified and to be solved. Recognizing the criminal activity patterns of a place is important in order to prevent it from happening. The crime solving agencies can do a better work if they have a good idea of the pattern of criminal activities that are happening in a particular area.

This can be done by using machine learning by employing different algorithms to find the patterns of the criminal activities in a particular area. This paper uses crime data set and predicts the types of crimes in a particular area which helps in speeding up the classification of criminal cases and proceed accordingly. This work uses the data of past 18 years that is collected from various trusted sources. Data pre-processing is as important as final prediction, this paper 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.

I. INTRODUCTION

Crimes are the significant threat to the humankind. There are many crimes that happens regular interval of time. Perhaps it is increasing and spreading at a fast and vast rate. Crimes happen from small village, town to big cities. Crimes are of different type – robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasing there is a need to solve the cases in a much faster way. But at present, the criminal cases that are pending in India are rapidly increasing with the number of crimes committed are increasing.

The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster. To solve a case based upon a particular data there should be a thorough investigation and analysis that is to be done internally. With the amount of crime data that is present in India currently the analysis and decision making of these criminal cases is too difficult for the officials. With the delay in decision making, prolonged cases and lack of equipment to handle the situation, the cases are destined to grow over the years.

II. LITERATURE SURVEY

[1] Mehmet Sait, and Mustafa Gök presented the criminal prediction for finding the most probable criminal of a particular offense incident when the suspected list of offenders are provided with the criminal data which is generated synthetically using Gaussian Mixture Model. The authors used Naïve Bayes performance. As a result of the comparison the authors achieved that the Naïve Bayes Classifier consumed less execution time and performs better with 78.05% accuracy. [2] Sivaranjani, S., S. Siva kumari, and M. Aasha Presented crime analysis for six cities of Tamilnadu, India by using clustering practices kmeans, DBSCAN and Agglomerative clustering for grouping the similar patterns to recognize offenses and the authors conclude that DBSCAN clustering performs better with precision 0.95, recall 0.91 and F measure 0.93 for grouping the similar patterns to identify crimes in for six cities of Tamilnadu, India. The authors used KNN practice to extract and predicts future offenses that will occur in the future in six cities of Tamilnadu, India which have possibility of low, medium and high offense occurrence by visualizing on google map.

[3] Emmanuel A., et al. Analyzed crime data by using support vector machine, naïve bayes, neural network and J48 and contrasts the techniques by using accuracy and execution time for predicting offense level as ‘Low’, ‘Medium’, and ‘High’. As a result of the contrast the authors conclude that the decision tree (J48) consumed less

execution time with 0.06 seconds and performs better with 100% accuracy for crime forecasting. [4] Yerpude P., et al. applied data mining practices from crime data for foreseeing features that affect the high or low crime rate in certain region. The authors used Random Forest, Naïve Bayes and Linear Regression for recognizing factors that affect the high crime rate and compared its performance. As a result of the comparison the authors conclude that the Random Forest performs better with 81.35% accuracy.

[5] Tahani A., et al. analyzed two different crime data using Decision Tree and Naïve Bayesian classifier to locate the most probable crime locations and their frequent occurrence time using Apriori Algorithm. The authors introduced what kind of offense might happen next in a specific place within a certain time and combining crimes' dataset with its demographics information to capture the issues that might disturb the safety of neighborhoods. As a result of the comparison the authors conclude that the Naïve Bayes performs better with 51% accuracy for Denver and 54% for Los Angeles for crime prediction.

[6] Sathyadevan, Shiju, and Surya Gangadharan proposed a model forecasts districts which have high likelihood of crime happening districts in particular day and Naïve Bayes predict with 90% accuracy for crime data analysis and foreseeing. Apriori Algorithm is used to find crime patterns that occur repeatedly in a particular district. The paper visualized high crime prone districts by GraphDB & data stored in mongoDB. [7] Tayebi MA, Ester M, Glässer U, Brantingham PL, Mohammad A. et al. presented CrimeTracer (which consists a criminal daily life, road network and whole places where criminal carry out a crime), random walk approach for crime location forecasting for urban area. The CrimeTracer is based on the probabilistic framework to model the features of known individual criminal within urban areas they are aware with, known as activity spaces.

[8] Zhang Q, Yuan P, Zhou Q, Yang Z. introduced enhanced multiple crime category called heat levels for predicting high crime intensity regions by dividing into five heat levels or category. The authors used LDA for feature reduction and KNN for prediction of high crime intensity regions. [9] Yu CH, Ward MW, Morabito M, Ding W presented predictive technique for areas that have a high density of residential burglaries and the probability that will happen in future. Primarily, a proper technique of organizing data is considered to store temporal/spatial information as well as combined numbers of offenses and offense-related actions by the police officer. Secondly, an ensemble supervised data mining algorithm is applied to achieve residential burglary prediction with better accuracy.

[10] Alkesh Bharati¹, Dr Sarvanaguru RA.K did a project that mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred. Using the concept of machine learning they have built a model using training data set that have undergone data cleaning and data transformation. The model predicts the type of crime with accuracy of 0.789.

[11] TeluguMaddileti, Vaddemani Sai Madhav, K V Sai Sashank, G. Shriphad Rao presented a paper where the prediction analysis done in the paper provides patterns of crime in a particular area by analysing and using certain machine learning algorithms. The relatively poor algorithm is Decision Tree classification which is having a low accuracy of 51.068%, and then the Logistic Regression with 78.9% accuracy. The author concludes that Random Forest is the best performed algorithm with an accuracy of 95.12%.

[12] Cesario, Cesario E, Catlett C, Talia D. They analyzed two various data sets for prediction of crime level category collected from different sources. The authors analyzed & compare Naïve Bayes & Neural network to predict crime level that will occur in particular location and Naïve Bayes classifier performs better with 90.2207% of accuracy.

III. EXISTING SYSTEM

Many researchers have gone through this problem regarding the criminal cases being unsolved for a long period. They proposed different crime prediction algorithms. In all these models the accuracy will surely vary depending on the data set and the features or attributes we select during data pre-processing. In Crime prediction done on the Mississippi crime data set where models like linear regression and Decision stump model are used gave a result of 83%, 88% and 67% respectively [1]. Although these accuracies of the predictions may vary accordingly because it is discovered that many machine learning algorithms are implemented on data sets consisting of different places having distinctive features, so predictions are changing in all cases

In the data set used is from kaggle.com and have selected models namely logistic regression, KNearest Neighbors (KNN), Decision tree classification, Bayesian methods. Data pre-processing is done by dropping the null values and filling the unknown values [2]. The accuracies obtained by using the above machine learning algorithms are

as follows 78.73% for KNN, 78.60% for decision tress classifier 31% for Support Vector Clustering (SVC) and 64.60% for Gaussian Naïve Bayes. The model is trained by cleaning and pre-processing the data and obtaining a best accuracy of 78.73%.

Disadvantages of existing system

By this methodology they had less accuracy in prediction

IV. PROPOSED SYSTEM

The proposed system is made on the basis of the research work that is done by going through various documentations that researchers and other worked on previously. On surveying previous works, Naive-Bayes, KNearest Neighbors, Decision Tree and Random Forest tend to give good accuracy so these models are used to predict crimes along with an additional boosting algorithm that is applied on some of the previously mentioned algorithms. The dataset used here is taken from data.world.com. The data set contains different types of crimes that being committed in India according to the state and year respectively. A CSV file containing different types of crimes is taken as input and the algorithms are tested for highest accuracy. The data pre-processing involves data cleaning, feature selection, dropping null values, data scaling by normalizing and standardizing.

After data pre-processing the data is free of null values which may affect the accuracy of the model significantly and feature selection is used to select only the required features that won't affect the accuracy of model. After data pre-processing the models chosen i.e., Naive-Bayes, KNearest Neighbors, Decision Tree, AdaBoost and Random Forest are trained by splitting the data into as train and test data. As the output required is a categorical value classification models are used here. Python language is used for the data prediction. Different visualizations are used to represent the relations and patterns in the data.

The advantages of this model are:

The accuracy of the model can be increased with the application of boosting algorithm.

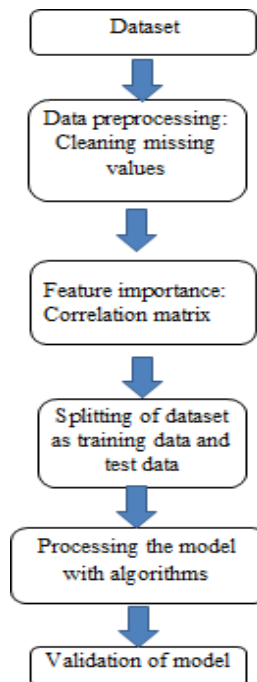
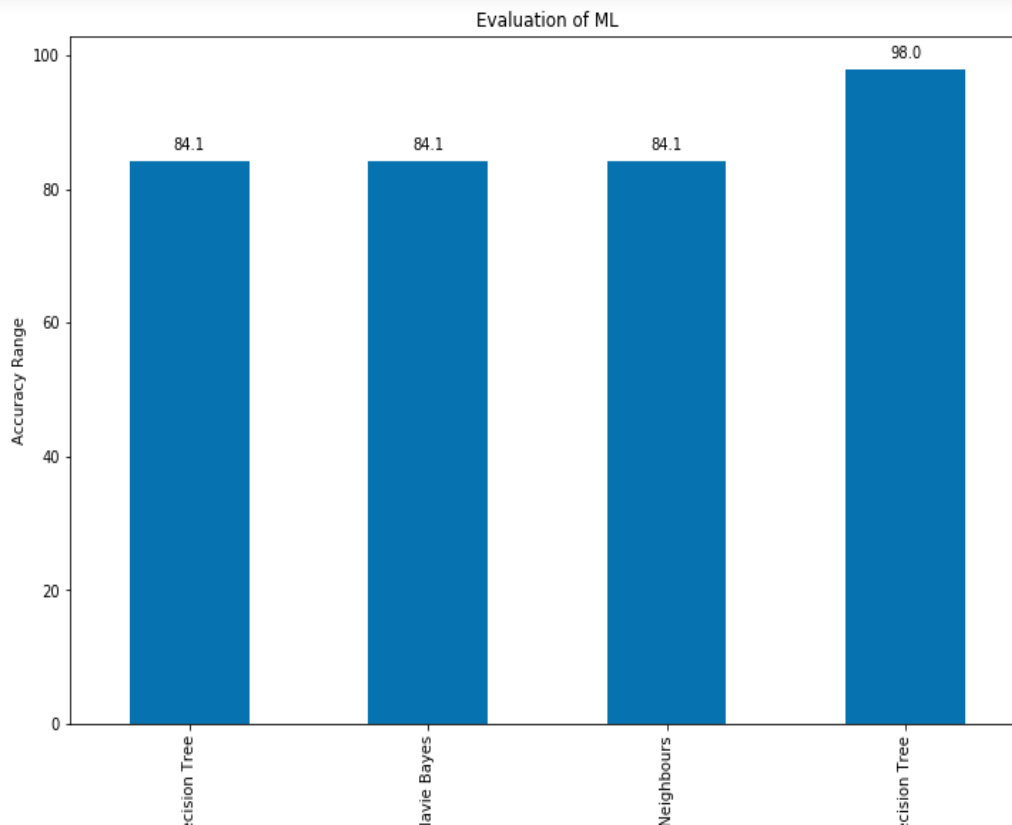


Fig: Architecture of proposed system

Result Comparison with Different Classifiers**V. CONCLUSION**

The prediction analysis done, provides patterns of crime in a particular area by analyzing and using certain machine learning algorithms. Among the used algorithms, Boosted Decision Tree Classification using adaboost is proven to be a better model for predicting the results which can be observed by comparing the accuracies with the other prediction algorithms. The relatively poor algorithm is boosted Naive-Bayes classification which is having a low accuracy of 62.71%. Hence, by this observation, it can be concluded that applying boosting algorithms do not always increase the accuracy. This project presents a crime data prediction by taking the types of crimes as input and the output is given using Jupyter notebook having python as a core language and python provide inbuilt libraries such as Pandas and Numpy through which the work will be completed faster and Scikit provides all the processes of how to use different libraries providing by the python. Results of prediction are different for different algorithms and the accuracy of Boosted Decision Tree Classifier found to be good with the accuracy of 97.99%.

VI. FUTURESCOPE

This project presented the techniques and methods that can be used to predict crime and help law agencies. The scope of using different methods for crime prediction and prevention can change the scenario of law enforcement agencies. Using a combination of ML and computer vision can substantially impact the overall functionality of law enforcement agencies. In the near future, by combining ML and computer vision, along with security equipment such as surveillance cameras and spotting scopes, a machine can learn the pattern of previous crimes, understand what crime actually is, and predict future crimes accurately without human intervention. A possible automation would be to create a system that can predict and anticipate the zones of crime hotspots in a city. Law enforcement agencies can be warned and prevent crime from occurring by implementing more surveillance within the prediction zone. This complete automation can overcome the drawbacks of the current system, and law enforcement agencies can depend more on these techniques in the near future. Designing a machine to anticipate and identify patterns of such crimes will be the starting point of our future study. Although the current systems have a large impact on crime prevention, this could be



the next big approach and bring about a revolutionary change in the crime rate, prediction, detection, and prevention, i.e., a “universal police officer”.

REFERENCES

- [1] Alkesh Bharati1, Dr Sarvanaguru R.A.K. Crime Prediction and Analysis Using Machine Learning. International Research Journal of Engineering and Technology (IRJET), Sep 2018.
- [2] Telugu Maddileti, Vaddemani Sai Madhav, K V Sai Sashank, G. Shriphad Rao. Crime Data Analysis using machine learning models. International Journal of Advanced Science and Technology 2020.
- [3] Cesario, Cesario E, Catlett C, Talia D. Forecasting Crimes Using Autoregressive Models. In Dependable, Autonomic and Secure Computing, 2016 IEEE 14th Intl C 2016 Aug 8 (pp. 795-802). IEEE.
- [4] Mehmet Sait, and Mustafa Gök. "Criminal prediction using Naive Bayes theory." Springer 28.9 (2016): 2581-2592.
- [5] Sivaranjani, S., S. Sivakumari, and M. Aasha. "Crime prediction and forecasting in Tamilnadu using clustering approaches." Emerging Technological Trends (ICETT), International Conference on. IEEE, 2016
- [6] Mahmud, Nafiz, et al. "Crimecast: A crime prediction and strategy direction service." 19th International Conference on. IEEE, 2016.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**[®]
cross **ref**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details