iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51705

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 11 Issue V May 2023- Available at www.ijraset.com

Crime Rate Prediction Using Naive Bayes and KNN Algorithm

Subhashini[1], Neha Kumari[2], Rimjhim Kumari[3], Mayank Kumar Singh[4], Mr. Sreenu Banoth[5]
[1, 2, 3, 4]*UG Student, Department of Computer Science and Engineering, IIMT College of Engineering, Greater Noida, Uttar Pradesh, India*
[5]*Assistant Professor, Department of Computer Science and Engineering, IIMT College of Engineering, Greater Noida, Uttar Pradesh, India*

Abstract: ***The criminal cases are increasing rapidly in our country due to which number of pending cases are also piling up. The continuous increase in the criminal cases is proving that it would be very difficult to be classified and to be solved. Recognizing the criminal activity patterns of a place is very important in order to prevent it from happening any other case. The crime solving agencies are doing a better work and can do a better work if they have a good knowledge of the pattern of criminal activities that are happening in a particular area or region. This problem can only be done by using machine learning any by using different algorithms to find the patterns of the criminal activities in a particular area or region. Data pre-processing is as important as final prediction and label encoding to clean and nourish the data. This research paper gives an efficient machine leaning model using different algorithms for predicting the next criminal case of the country.***
Keywords: Machine Learning, Data Mining, KNN Algorithm, Bayes Theorem.

I. INTRODUCTION

To break a case grounded upon a particular data there should be a thorough disquisition and analysis that's to be done internally. With the quantum of crime data that's present in India presently the analysis and decision timber of these felonious cases are too delicate for the officers. Relating this a major problem this paper concentrates on creating a result for the decision timber of crime that's committed. the vehicle starts driving on its own. An independent driving vehicle performs colorful conduct to arrive at its destination, repeating the way of recognition, judgment and control on its own. In addition, represent the most crime rates in larger metropolises, crime reducing is getting one of the most important social issues in enormous metropolitan areas as it affects people security issues, youth growth and person social- profitable status. Crime rate cast is a scheme that uses different algorithms to determine the crime rate grounded on previous information. For our daily purposes we've to go to numerous places every day and numerous times in our daily lives we face multiple security issues similar as kidnapping, hijacking, importunity, etc. In addition, represent the most crime rates in larger metropolises, crime reducing is getting one of the most important social issues in enormous metropolitan areas as it affects people security issues, youth growth and person socio-profitable status. This exploration introduces the design and prosecution of a strategy grounded on once crime data and analyses the crime rate in once areas at distinct moments; for this work, we use primary data those are collected from the people grounded on their former crime problem. In our train information collection, we use different algorithms to figure out the loftiest perfection between the KNN algorithms that provides the topmost perfection. Crime prognostications can be made through both qualitative and quantitative styles. Qualitative approaches to soothsaying crime, as environmental scanning, script jotting, are useful in relating the unborn nature of felonious exertion. In discrepancy, distinguishable styles are used to prognosticate the unborn compass of crime and more specifically, crime rates a common system for develop vaticinations is to systems periodic crime rate trends developed through time series models. This approach also involves relating once crime trends with factors that will impact the unborn compass of crime.

II. RELATED WORK

The researchers are proposing their work related to personal crime rate prediction.
McClendon, Lawrence, and Natarajan Meghanathan proposed that data mining and machine literacy have come a vital part of crime discovery and forestalled. In this exploration, we use WECA, an open-source data mining software, to conduct a relative study between the violent crime patterns from the Communities and Crime unregularized Dataset handed by the University of California-Irvine depository and factual crime statistical data for the state of Mississippi that has been handed byneighborhoodscout.com.

©IJRASET: All Rights are Reserved

We enforced the Linear Retrogression, cumulative Retrogression, and Decision Refuse algorithms using the same finite set of features, on the Communities and Crime Dataset. Overall, the direct retrogression algorithm performed the stylish among the three named algorithms. The compass of this design is to prove how effective and accurate the machine learning algorithms used in data mining analysis can be at prognosticating violent crime patterns.

Wang, Tong, etal. proposed that we introduce a novel, robust data- driven regularization strategy called Adaptive Regularized Boosting (AR- Boost), motivated by a desire to reduce overfitting. We replace AdaBoost's hard fringe with a formalized soft fringe that trades- off between a larger fringe, at the expenditure of misclassification crimes. Minimizing this homogenized exponential loss results in a boosting algorithm that relaxes the weak knowledge supposition further it can use classifiers with error lower than 12. We decide bounds for training and generality crimes, and relate them to AdaBoost. ultimately, we show empirical results on standard data that establish the robustness of our approach and bettered performance overall. 1 prolusion Boosting is a popular system for perfecting the delicacy of a classifier. In particular, AdaBoost is considered the most popular form of boosting and it has been shown to meliorate the performance of base learners both theoretically and empirically. The pivotal idea behind AdaBoost is that it constructs a strong classifier using a set of weak classifiers.

Hardi, M. Patel, Ripal Patel proposed that data mining is used considerably in terms of analysis and disquisition of patterns for circumstance of different crime. Data mining can be used to model crime discovery problems. Crimes are a social nuisance and bring our society dearly in several ways. Our end is to prognosticate the crime and position in which specific types of crime will do. Our approach is grounded on novelettish analysis by applying wordbook- grounded system. We've to use then twitter data set. We're prognosticating unborn crime in particular area (Illinois, Chicago).

Jyoti Agarwal, Renuka Nagpal and Rajni Sehgal proposed that in moment's world security is an aspect which is given advanced priority by all political and government worldwide and aiming to reduce crime frequency. As data mining is the applicable field to apply on high volume crime dataset and knowledge gained from data mining approaches will be useful and support police force. So, in this paper crime analysis is done by performing k- means clustering on crime dataset using rapid-fire- fire miner tool.

III. ARCHITECTURE DESIGN

A. Import Libraries

In this module different libraries are imported, which are useful for data processing, visualization and vaticination etc.

B. Data Processing

In this module data undergoes pre-processing, like cleaning of dataset and removing null and unwanted values etc.

C. Visualization of crime against rape

We've different times of crime data but originally, we're taking rape cases. In this module data undergoes Visualization of crime against rape.

D. Model Generation

In this module model is figured.

E. Figure Naïve Bayes Gaussian Classifier

In this module Naïve Bayes Gaussian Classifier is used to data analysis.

F. Figure Decision Tree Classifier

In this module Decision Tree Classifier is used to data analysis.

G. Figure KNN Classifier

In this module KNN Classifier is used to data analysis

H. Figure Boosted Decision Tree Classifier

In this module Boosted Decision Tree is used to data analysis

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

I. Delicacy Comparison

In this module delicacy comparison is shown. Delicacy is compared among the different kinds of classifiers used and we noticed that using Boosted Decision Tree Classifier High Accuracy is observed.

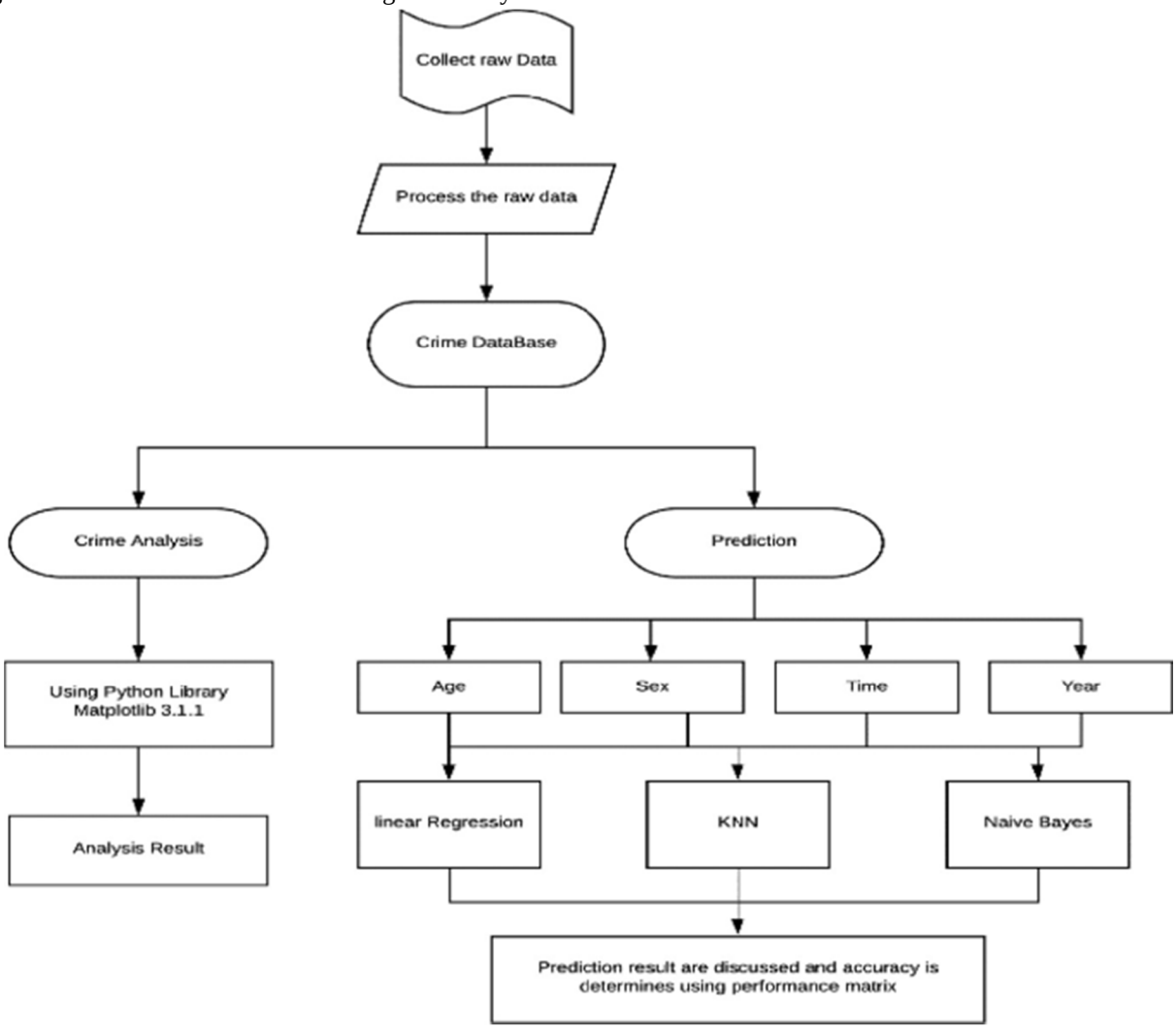


Fig. 1 ARCHITECTURE DIAGRAM

IV. ALGORITHM

A. Naïve Bayes Classifier

Naïve Bayes is a supervised machine learning algorithm used to break bracket problems. It's the simplest bracket algorithm. This algorithm is grounded on Bayes Theorem. The formula of Baye's theorem is-

$$P(A|B) = P(B|A) * P(A)/ P(B).$$

This algorithm fluently detects the stress and gives the result in double form.

B. K-Nearest Neighbors

K- nearest neighbors is used when the target variable must be classified in further than two classes. In this dataset, there are three classes of target variable perpetrator coitus joker, womanish, and unknown. also, three orders of youthful, old and sprat are defined in age. To classify these target variables, K- nearest neighbors Classifier is used.

$$D(a,b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

©IJRASET: All Rights are Reserved

V. CONCLUSION

It's clear that introductory details of felonious conditioning in a neighborhood contain pointers that will be employed by machine literacy agents to classify a felonious exertion given a position and date. The training agent suffers from imbalanced orders of the dataset, it had been ready to overcome the problem by oversampling and under-slice the dataset. This paper presents a crime data vaticination by taking the types of crimes as input and producing those crimes which are committed as affair using Jupyter tablet. Results of vaticination are different for different algorithms and the delicacy of Boosted Decision Tree Classifier set up to be good with the delicacy of95.122.

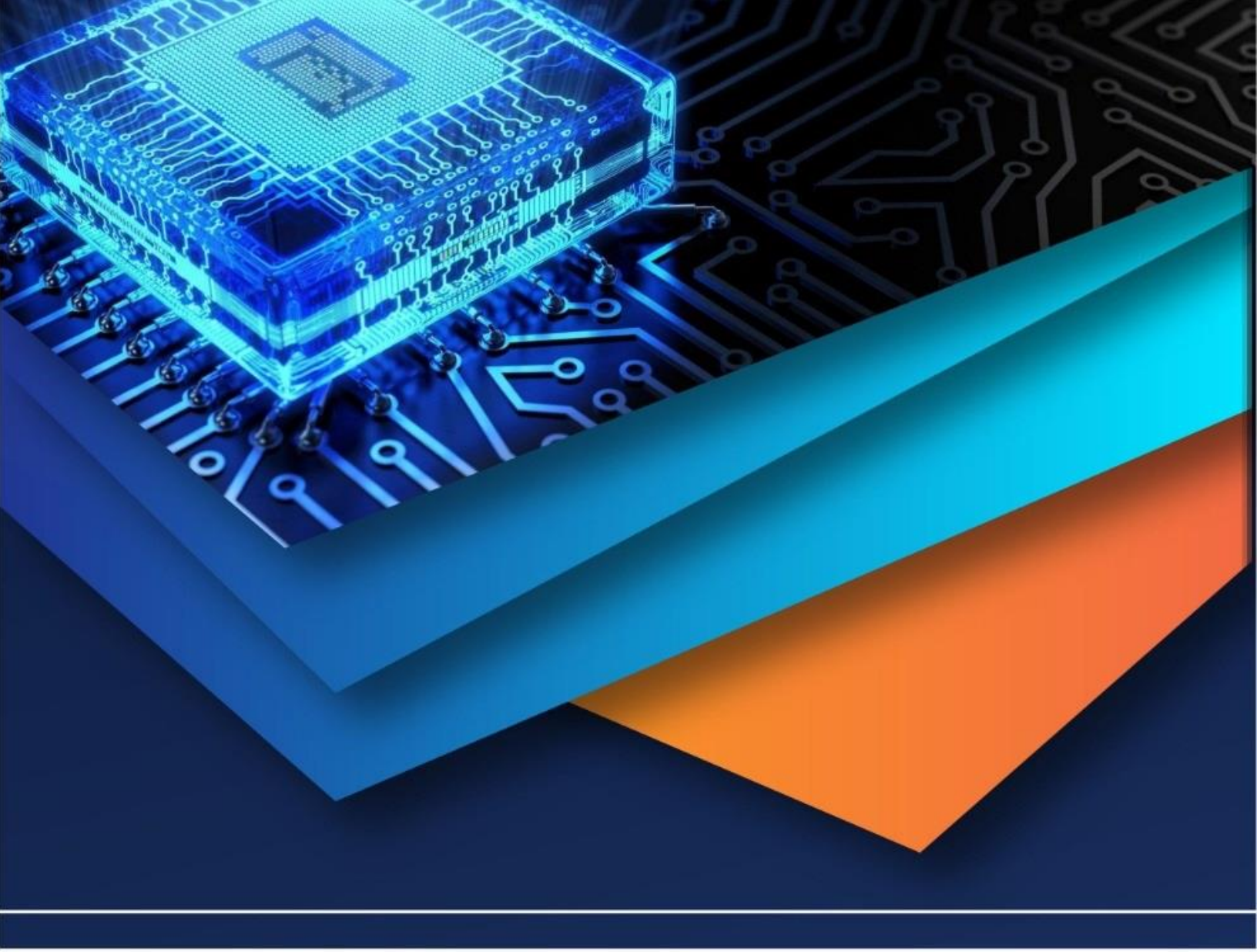
VI. FUTURE ENHANCEMENT

From the encouraging results, we believe that crime data mining has a promising future for adding the effectiveness and effectiveness of felonious and intelligence analysis. Visual and intuitive felonious and intelligence disquisition ways can be developed for crime pattern. As we've applied clustering fashion of data mining for crime- analysis we can also perform other ways of data booby-trapping similar as bracket. Also, we can perform analysis on colorful dataset similar as enterprise check dataset, poverty dataset, aid effectiveness dataset, etc.

REFERENCES

[1] McClendon, Lawrence, and Natarajan Meghanathan. Using machine knowledge algorithms to anatomize crime data. Machine knowledge and operations An International Journal (MLAIJ)2.1 (2015), 1- 12.
[2] McClendon,L., & Meghanathan,N. (2015). Machine knowledge and operations An International Journal (MLAIJ), 2 (1), 1- 12.
[3] Sathyadevan, Shiju. Crime analysis and prophecy using data mining. 2014 First International Conference on Networks & Soft Computing (ICNSC2014). IEEE, 2014.
[4] Sathyadevan, S (2014, August). In 2014 First International Conference on Networks & Soft Computing (ICNSC2014) pp. 406- 412). IEEE.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)