Crime data analysis and prediction using ANCalgorithm

[1] Dr.S.Shanmuga Priya, [2] Ramya Srinivasan, [3] Prisila Johnpeter, [4] Hemalatha Selvaraj,

[5] Dr.L.Mary Gladence

[1] Associate Professor, Department of Computer Science and Engineering M.I.E.T. Engineering college,Trichy

[2, 3, 4] UG Student, Department of Computer Science and Engineering, M.I.E.T. Engineering college, Trichy

[5] Associate Professor, Sathyabama Institute of Science and Technology,Chennai

[1] priya501@gmail.com [2] nalluramya23@gmail.com [3] ammuprisila1@gmail.com [4] hemamaluy1618@gmail.com, lgladence@gmail.com

Abstract— A crime is an illegal activity that affects a community and affects the peace of the society . These illegal activities are condemned by the court of justice through the corresponding punishment written by the justice. Criminal offenses are often dealt by the state government or the commonwealth by the action of the court. In recent times many authors have tried to predict the offenses for the future by using the latest algorithms .The study is useful for police officers to reduce the crime rates in our society.

Index Terms— Ada Booster, Apriori , Big Data, C5.0,Naives bayes

I. INTRODUCTION

Big data is a large amount of data that occupies a lot of storage yet increases with time. Criminalism is an illegal offense that badly affects the peace of the society and also the happiness of the citizens. The crimes and criminal acts can cause loss of money to both public and private sectors. Public safety is a factor for securing the environment when people travel to other places. In day to day life various types of crimes are associated with consequences. On the whole crime takes place due to several reasons or motives, behavior and human nature etc. Crime forecasting has earned familiarity in the last few years because it helps in investigating a crime and also to handle a crime computational. There is a bid requirement for an efficient predictive algorithm which directs police keeps towards criminals [4:54 pm, 22/07/2022] Amma Kutty: In day to day life various types of offenses are associated with some relations .on the whole crime offense takes place due to several reasons or motives behavior and human nature Furthermore crime forecast have now developed to forecast the geographical location of the offense. Analyzing this kind of data will effectively keep track of the events happening presently, identifying some similarities from the incident and also useful to make effective decisions.

Analyzing these crime offenses enables the investigation officers to get a clear view of the crime occurrence and also to prevent the crime occurrence. The accurate forecast of crime

analysis creates many challenges and difficulties.The aim is to analyze the offenses and forecast for the future all over India . Different machine learning algorithms are chosen for the forecast.

Many researchers have indulged themselves in the crime forecast for future generations such as naive Bayes, k-means clustering technique, c4.5 c5.0 random forest, Apriori algorithm, lstm method ,Arima model etc. Analyzing this kind of data will effectively keep track of the events happening presently in the society, identifying some similarities from incidents and also useful in making effective decisions. The criminal offenses affect both the wealth and also the health of the society.public safety is the major factor that has more weightage when mentioning the crime occurrence by the police officers.

EASE OF USE

A.   Literature Review

The latest literature concerning offense knowledge analysis may be categorized by several analyses. it additionally counsels that crime analysis and prediction square measure supported new kinds of knowledge taken from online forms The highlights of the study are the classification ,prediction and forecast of crime offenses. Different kinds of views of crime forecast are analyzed by numerous ways,most typically the forecast is done on the given computer file

.National Crime Records Bureau NCRB is the official website of India.With this goal there are several numerous studies in the previous years

A. Crime Data Analysis using Formal Concepts Analysis Algorithm associated with fuzzy conception lattice for association rules mining. projected a associate degree methodology for analyzing knowledge in relative to the periodicities, provided solutions to unravel human trafficking problems exploitation fuzzy incidence graphs. Discover a number of the crime patterns in unipolar house exploitation, the entropy theory and granular computing. Introduced many ways to alter fuzzy attributes on the far side unipolar and multipolar groups with their graphical structure. Formal concept analysis is one of the most important aspects of the crime forecast. A unique feature of this is the combination of 3 Elements like the Invention and reasoning with dependencies and also the graded indexed. Discover the number of crime patterns in unipolar house exploitation 8th entropy theory and other granular computing . United nations organizations Age would like to start ot analysis during this space. The methodology of different authors is delineated beside the conclusion and analysis gapThe inputs in FCA are diagrammatic within the style of Table that is thought as a cross table. The objects square measure diagrammatically by the rows of the table, and also the columns represent the attributes. For a proper context, component p from P square measure referred to as objects and component letter from letter square measure referred to as attributes.

B. Crime Prediction and Analysis

Criminal offense is the most ever growing problem of society . Crime patterns measure

perpetually as a result of that it's tough to clarify the behavior in the patterns . The enforcement officers collect the crime data knowledge into technologies however incidence of any Crimes

Is of course difficult to predict and from the full research of Crimes it has been found numerous factors like economic conditions unemployment affect the crime rate

But the difficulty in the research is the gathering of trueful crime rate accurately . A strong prediction system is required to analyze the current scenario

The study focused on the major algorithm like kn n classifier, support vector machine svm are used to predict the offense and artificial neural networks is used as the backbone for the study

With the tremendous increase in crime category analysis of crime is most adequate .crime analysis generally have default step is thaT to maintain the peace of the society as it is. The study also adds the additional information regarding the crime as it may be foretold that the place and time of the offense is likely to be unmatched.

Classification of crime includes the Removal of the unwanted details about the crime

Data Preprocessing

The crime data used for the analysis and forecast consists of the list of crime records every where Asian nation and surrounding.We have non inheritable crime knowledge set from the genuine government web site referred to as NCRB National Crime Record Bureau.The information set from the NCRB provides data from 2001 to 2020 with millions of population density. The accumulated information set from on-line repositories typically contains irrelevant info. The overall review of the obtained knowledge contains tons of noise, inconsistencies, bugs bungles and conjointly some quite unwanted knowledge. To stop this type of issue, choice of meaning and genuine knowledge is important to avoid data and garbage values.[8]. And delete the unwanted garbage values and also the missing and therefore convert the data set to a structure and also to a helpful knowledge to realize information handling. Therefore converts the impracticable knowledge to a possible and helpful knowledge to realize info handling [9]. To boot collection the additional mind-blogging framework is often needed keeping visible the developing rates of knowledge in business analytics, analysis areas, science. pre process firmly assures data coming up with, improvement institutionalism, and alter data assignments, thereby decreasing the varied style of the data bodily process of unwanted parts from the info through assurance of parts , discretization frames and finally it insists on developing applied mathematics vital knowledge to create correct crime knowledge prediction[10]. For implementation Python language for the info transformation. the ultimate attributes taken for the analysis square measure form of crime, crime location, year of the crime, postcode and conjointly police district. thus the info set is split into a pair of sections like check sets (30%) and coaching sets(70%) [11].Finally a specific quantity of instances are taken when the preprocessing step .Accurate, precision ,recall, and conjointly the f1 scores square measure the most parameters used for analysis of performance during this study [12][14]

4. Prediction and Forecasting

Forecast And analysis of crime has got a lot of changes in recent times through the development of technologies and It. Crime forecast is done through the crime data that is got for the past few years and analyzed   the accuracy and found . On the other hand, prognostication is done towards the long run of crime trends and future crime hotspots. A fast review of all the criminal activities may be achieved by the investigation officers through the on the market software system packages[2]. Whereas deep learning approaches might guarantee optimum prediction. Different methods were employed in the review papers like k means that clump, algorithm, R tool KNN formula,Naives Bayes formula, ARIMA, LSTM model and even a lot of analysis teams around the world have recently used totally different methodologies like supervised models and additionally unsupervised models like ARIMA model, motor vehicle Regression, regression, Prophet model and etc. so C.5 formula is employed for higher cognitive process mistreatment tree classifier, Apriori formula for calculating the association between one field with different and enzyme booster for increasing the potency of the prediction of crime on the given knowledge set.

5. Data Analysis and Visualization

The data set that we've used for the prediction analysis are globally on the market that covers Bharat through the official website of the presidency of India-NCRB- (National Crime Record Bureau). The crime knowledge contains to 2165954 crime offenses from 01/01/2001 to 6/10/2020

1. C5.0 Algorithm

The C5.0 model is the call tree formula with most gain. every bit is divided by the one split is then split once again, sometimes supporting a distinct field,until it reaches a condition that it can be divided anymore. Finally, the least splits are examined once more, and therefore the rest are aloof from the tree. The C5.0 node will predict solely a selected categorical target. Once analyzing knowledge with categorical fields, the node is probably going to cluster classes along. C5.0 will manufacture 2 differing types of models. a call tree may be an uncomplicated description of the splits found by the formula. every leaf node describes a selected set of knowledge, and every case within the coaching knowledge belongs to precisely one terminal node of the tree. In different words, just one prediction is going to be on the market for a selected set of knowledge in the call tree.The foremost necessary distinction between a call tree and a rule set, if quite one rule is applied to any explicit record, or not even one rule might apply. If too many rules are applied to an equivalent record, every rule is going to be weighted to a & quot; vote & quot; supports the arrogance with the rule, and therefore the final result is determined by a combination of all the weighted votes that's applied to the information sets given . If no rule is applied, a default basic prediction is appointed to the information set.

2. Apriori Algorithm

Apriori Algorithm sets relevant association rules between the given attributes. Apriori algorithm operates on a huge number of transactions that are present in a database. Apriori

algorithm helps the customers to buy products in a user friendly manner and increases the sales of a particular store. There are 2 types of components available in apriori algorithm . Support. Confidence and Lift .Support refers to the familiarism and also necessity of a product. Confidence refers to the possibility that the clients will buy both products. Lift refers to the increase in the ratio of the sale

3. Naives Bayes

Identifies the pattern of a given data set that utilizes. Bayesrules is the foundation of Naives Bayes with a strong presumption that the attributes are platform independent, When this independence is violated naive Bayes always delivers competitive classification with maximum accuracy. Coupled with the computational efficiency of naives bayes and many other desirable features leads to identify thepattern in sample data to approximate prediction of the postprobability P(y | x) of each class y,for the given object x. Naive   Bayes   properties   include   Computingefficiency,processing time is less with respect to both the training examples, i.e. the number of attributes, andclassification time is linear.
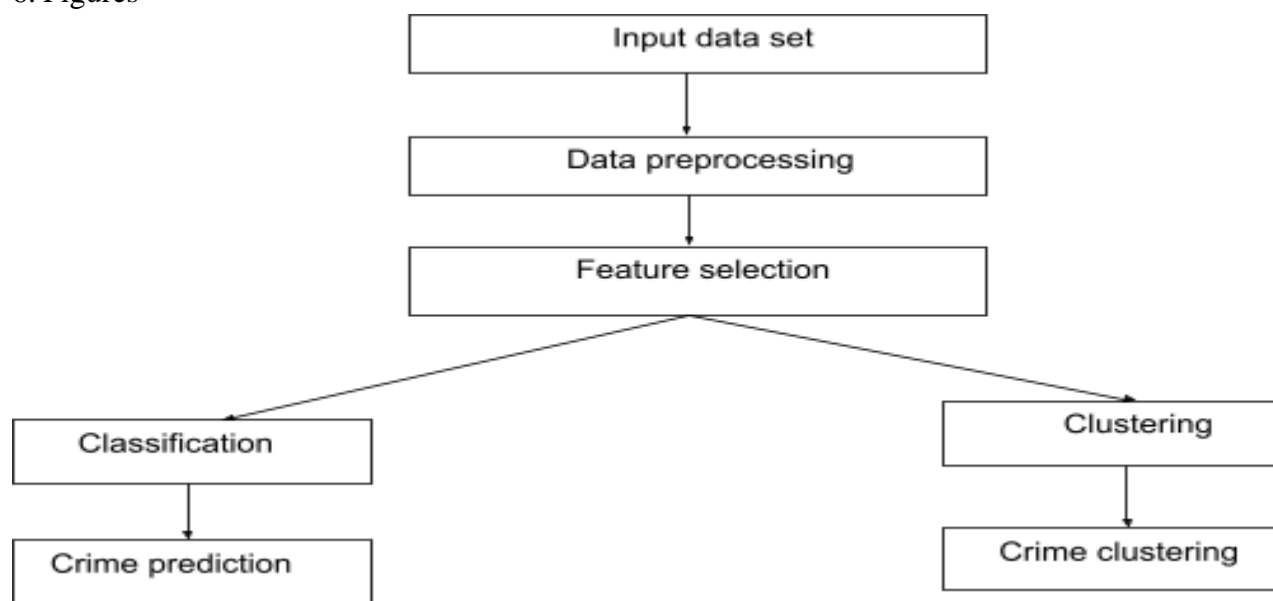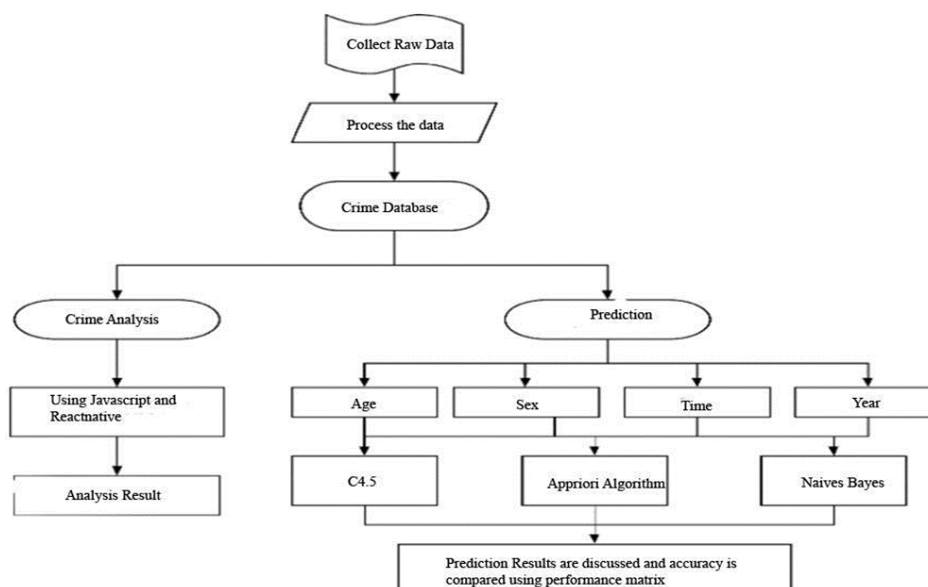
6. Figures

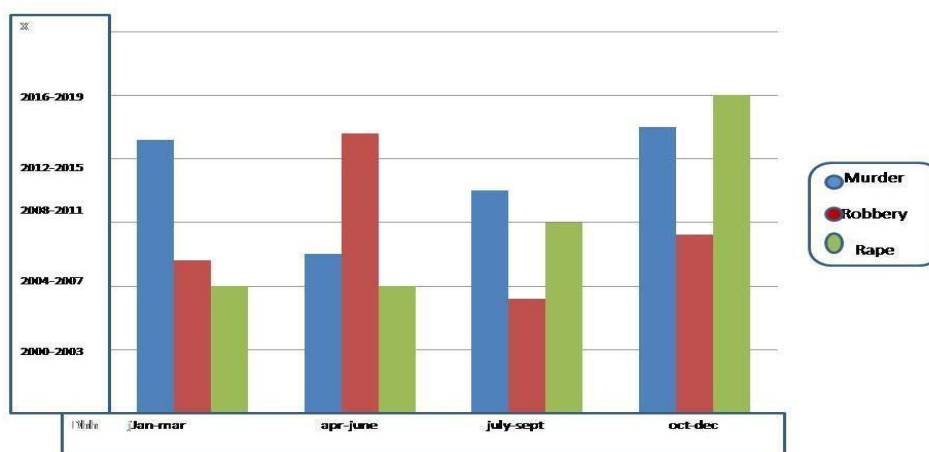Fig.1 proposed system

Fig.2 Empirical analysis of crime

fig 3. Crimes based on the years

6. Conclusion and Future work

Crimes are the biggest threat to human resources and also an important factor for decreasing the growth of our nation. Investigation authorities require a computational forecasting system to improve crime analysis and also to prevent the occurrence of an offense. Herein we have achieved an improved and standardized prediction model with utmost accuracy by indulging multiple techniques

The exploratory crime data analysis helped for a clear visualization of the crime patterns, crime rates across India and also the best algorithm to analyze the crime data sets.

Future Enhancement of this study will be crime profiling, which can be done to investigate the crime in a easy method More crime classification and prediction algorithms can be implemented to increase crime prediction accuracy and to enhance overall performance .It is also a useful extension for considering the income information for neighborhood in order to see whether they are having any relationship with the crime. By applying this methodology

all over the world with corresponding visual data for different crime data sets will create a crimeless peaceful society.

REFERENCES

1. prerna kapoor, Aswani Kumar, Premkumar Singh "1 Crime data set analysis using formal concepts Analysis (FCA)" April-2020
2. Wajiha Safat, Sohail Asghar,(member ieee) and Saira Andleeb gillani "Empirical analysis for crime prediction [3] Saravanan parthasarathy, Arun raj, Selva prabha "Survey on crime analysis and prediction using machine learning techniques and data mining techniques"
3. Prema Sundari and Yamini "A violent crime analysis using fuzzy c means clustering algorithm"
4. eugenio Cesairum, Domenica Talka, Charles E. Catlet "Forecasting crimes using auto regressive model"- 2019
5. Nikitha patil, Subhanji patil, Somanth "Crime analysis and prediction using data mining techniques" – 2019
6. Lawrenz E colsen, Marcus Felson "Social change and crime rate trends A routine activity Approach" – 1979
7. Pratibha, Akanishka Gahalot, Uprant "Crime prediction and analysis"- 2020
8. Bhajneet kaur, laxmi, vinay "Crime against women: analysis and prediction using data mining techniques" – 2019
9. S.Prabhakaran, Shilpa mithra "Survey of analysis of crime detection techniques using data mining and machine learning approach" – 2018
10. xiangyu zhao, Jiliang tang "Exploring transfer learning for Crime prediction"-201
11. Agaba joab ezra, Erio Crucesia, Binja Amoo Brenda "Crime analysis and prediction using data mining"- 2015
12. Gayathiri, srividhya, Srinivas "Crime data analysis and prediction"– 2018
13. Gladence, L.M., Anu, V.M., Revathy, S. et al. Security management in smart home environment. Soft Comput (2021).