A NOVEL APPROACH FOR CRIME PREVENTION ANALYSIS WITH MACHINE LEARNING

[1]ESLAVATH SUPRAJA(M.Tech)[2]P. R. Rajesh Kumar M. Tech, (PhD)

[1,2]SRI KRISHNADEVARAYA UNIVERSITY COLLEGE OF ENGINEERING AND TECHNOLOGY, SK UNIVERSITY, ANANTAPUR PIN CODE:515003

[1]suprajaeslavath123@gmail.com, [2]prrajeshcse@skucet.ac.in

Abstract

In this era of recent times, crime has become an evident way of making people and society under trouble. An increasing crime factor leads to an imbalance in the constituency of a country. In order to analyse and have a response ahead this type of criminal activities, it is necessary to understand the crime patterns. This study imposes one such crime pattern analysis by using crime data obtained from Kaggle open source which in turn used for the prediction of most recently occurring crimes. The major aspect of this project is to estimate which type of crime contributes the most along with time period and location where it has happened. Some machine learning algorithms such as Naïve Bayes is implied in this work in order to classify among various crime patterns and the accuracy achieved was comparatively high when compared to precomposed works. Keywords: Crime, Analyse, Crime patterns, Kaggle, Estimate, Naïve Bayes, Accuracy.

Keywords : svm, crime,type, es.

1. INTRODUCTION

About Machine Learning

Computer algorithms that learn better on their own are the focus of machine learning (ML). Artificial intelligence (AI) incorporates it. To pursue forecasts or decisions without being unequivocally shown to do as such, AI calculations build a numerical model utilizing test information, alluded to as "preparing information". Email sifting and PC vision are only two examples of the many uses for machine learning algorithms, which are useful in situations where it would be unreasonable or difficult to utilize customary calculations to do the expected undertakings.

The field of computational measurements, which fixates on the utilization of PCs to

make forecasts, is closely connected to machine learning. Machine learning draws theory, methodology, and application fields from mathematical optimisation. An related area of research, data mining is concerned with the use of unaided learning for exploratory information investigation. AI is here and there called prescient investigation when it is utilized to settle different business challenges.

Machine learning is the process by which computers learn to carry out tasks automatically, without any human intervention. Computers are taught to do certain jobs by analysing and interpreting data that is given to them. It is feasible to programme algorithms that instruct computers to carry out all the necessary procedures to solve basic problems; in these cases, the computers do not need to learn anything. It might be hard for people to physically foster the calculations expected for progressively complex positions. It very well might be moreefficient in the long run to let the computer figure out the algorithm on its own rather than to have humans code it step by step.

When there isn't yet a perfectly adequate algorithm for a certain task, researchers in the field of AI utilize various proceduresto teach computers how to do the job. One way to deal with situations when there are many

possible responses is to mark some of the right ones as legitimate. The data may be sent into the computer's training algorithms, which can then be fine-tuned to provide more accurate results. As an example, the MNIST dataset is frequently used to prepare frameworks for computerized character acknowledgment assignments.

Machine learning approaches

There were three main groups into which early machine learning techniques were grouped, according on the sort of "sign" or "input" that the learning frameworkhad access to. Here they were:

Computers may become familiar with a basic principle that guides contributions to yields utilizing directed realizing, which includes a"teacher" presenting the computer with sample inputs and their intended outcomes.

Without labels to guide it, a learning system may discover patterns in unsupervised data. Discovering hidden patterns in data is one example of an end aim of unsupervised learning, whereas feature learning is another.

A computer programme engages in reinforcement learning when it encounters a constantly changing environment that requires it to complete a specific task (like driving a vehicle or contending in a game).

As it travels through its concern space, it receives feedback that is similar to rewards, which it strives to maximise.

AI frameworks frequently utilize more than one of thesethree types of methodologies; also, new methods have emerged that do not cleanly fall into any of these categories. Examples include meta-learning, dimensionality reduction, and topic modelling. [8] In 2020, deep learning has surpassed all other approaches in machine learning as the method of choice for most active projects.

ABOUT THE PROJECT

A key thread that is thought to be growing rather intensely is crime, which has just emerged. A statement is considered criminal if it is both very offensive and in violation of government laws. Research into the several subfields of criminology as well as the art of pattern recognition is necessary for crime pattern analysis. It takes a lot of effort and time for the government to utilize innovation to control a portion of theseillegal operations.

Therefore, in order to forecast the kind and frequency of crimes, machine learning methods and their associated data are necessary. It makes use of preexisting crime data to forecast the kind and frequency of crimes based on time and place. Numerous studies have been conducted by researchers to analyse crime trends and their connections in a particular area. It is now simpler to categorise crime trends in some of the examined regions. Officials will actually want to settle gives all the more rapidly subsequently. Using a dataset derived from the open-source Kaggle platform, this method takes into account not just the time and location of an event, but also a number of other characteristics. We included a categorization algorithm that aids in identifying the nature of the crime and the locations where it is most prevalent at a given time and day. Here, we suggest using machine learning methods to scour the provided geographical and temporal data for crime patterns that satisfy our criteria.

OBJECTIVE OF PROJECT

Applying machine learning algorithms to the provided geographical and temporal data in order to identify criminal patterns and their associated categories is the primary goal of this study.

2. LITERATURE SURVEY

examination of criminal activity using machine learning

Predicting criminal behaviour using machine learning is the focus of this research. This study employs two distinct

data-processing methodologies to examine Vancouver crime statistics over the last fifteen years. A wrongdoing expectation precision going from 39% to 44% is accomplished in Vancouver utilizing AI prescient models, specifically K-closest neighbor and helped choice tree.

An Overview of Data Mining for Crime Analysis and Prediction

Data mining is the process of exploring and analysing huge current datasets to discover new information that can be crucial to the company. It is anticipated that additional information will be extracted from preexisting databases. There have been a lot of data mining techniques to analysis and prediction. However, the discipline of criminology has made very little progress. Only a small number of people have made an attempt to compare the data produced by each of these methods. There is a wealth of data stored by police stations and other comparable criminal justice organisations that may be mined for insights about criminal behaviour and trends in society. Additionally, the offenders might be forecasted using the crime statistics. The primary objective of this study is to conduct an overview of the methods used for criminal identification that make use of supervised and unsupervised learning. An overview of crime analysis and prediction utilising various data mining approaches is provided in this study.

Methods for Detecting and Predicting Criminal Events: A Comprehensive Review

There has to be a system in place that can identify and forecast crimes in real time since the crime rate is going up. Data mining methods including affiliation rule mining, k-implies grouping, choice trees, and guileless bayes, as well as AI procedures like profound brain organizations and counterfeit brain organizations,are the focus of this survey. The goal is to examine these methods in order to discover and forecast crimes.

This study found that crime is not evenly distributed throughout metropolitan landscapes but rather focuses in certain places, and that pre-processing becomes an essential activity when the dataset occurrences have countless missing qualities. Accordingly, using post-processing to forecast crime hotspots is a crucial step towards reducing crime rates.

Analysis and Prediction of Criminal Patterns using Machine Learning

In addition to being an annoyance to society, crimes have significant monetary and non-monetary costs. The investment in research

that yields faster crime solving results will be more than recouped. Only about 10% of offenders are really responsible for 50% of all crimes. For training purposes, the system is fed data from official Indian crime databases that detail a wide range of offences, including rape, kidnapping, dacoits, robbery, burglary, and murder. An analysis was conducted using data from the Indian statistics, which covers a range of crimes committed between 2001 and 2014. A regression model was then developed to forecast future crime rates in different states. Crime data have been subjected to supervised, semi-supervised, and unsupervised learning techniques in an effort to find new information and improve the accuracy of crime predictions. When it comes to reducing crime, this effort will be useful to the local police stations.

3. DEEP NEURAL NETWORK

The first step in pre-processing a dataset is to identify and delete duplicate values and characteristics. This is done with datasets that have been received from free sources. Both the identification of criminal trends and the extraction of characteristics from massive datasets have made use of decision trees. It lays the groundwork for further categorization. Using a Deep Neural Network, we are able to extract features that

pertain to the identified crime patterns. Both the training and test values' performance is computed using the prediction. Predicting when criminal acts of any kind are likely to occur in the future allows law enforcement to address these issues as soon as possible.

3.1 LIMITATIONS

The classifier in previous studies employs categorical values, which leads to a biassed result for the nominal qualities with higher value, which explains why the accuracy is poor. Using the categorization approaches in areas with genuine valued qualities and inadequate data is not a good idea. It is necessary to give an appropriate value to the classifier as its value needs tuning.

4. AI BASED TECHNIQUES

Utilizing the AI procedures of channel and covering, the obtained information is first pre-handled to dispense with copy and immaterial qualities. Thedata has been cleaned up since it decreases the dimensionality as well. After then, the data is separated even more. There is a test informational collection and a preparation informational collection. Both the preparation and testing datasets are utilized to prepare the model.After that, mapping comes next. In order to facilitate categorization, the following fields are translated to integers: crime category, year,

month, time, date, and location. From the get go, Gullible Bayes is utilized to analyze the autonomous impacts between the qualities.To categorise the retrieved independent characteristics, Bernoulli Naïve Bayes is used. You may examine the crime incidence at a certain time and area by labelling the criminal aspects. The most widely recognized violations, along with their topographical and transient events, are at last revealed. By deciding the precision rate, one might learn the presentation of the expectation model. Python, an internet based compiler for information investigation and AI models, was utilized to make theprediction model and execute it on the Colab.

FEATURES OF AI BASED TECHNIQUE

Since the majority of the included qualities are time and place dependent, the suggested method is ideal for crime pattern recognition. What's more, it gets rid of the issue of analysing the independent effects of the attributes. Since it takes into consideration actual and nominal values and also worries the area with inadequate data, the ideal value initialization is unnecessary. In comparison to other machine learning prediction models, the accuracy has been pretty good.
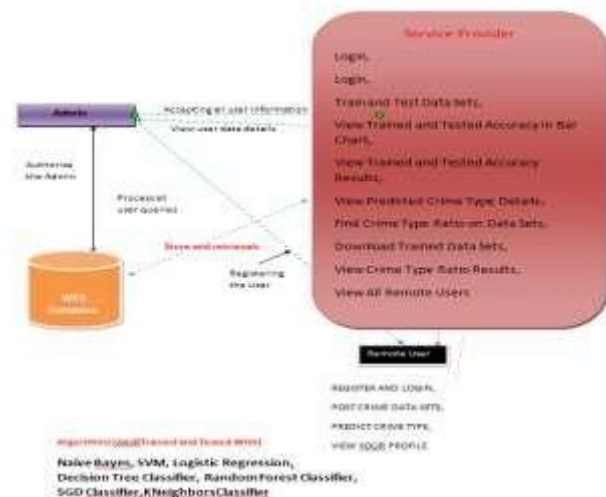
5. ARCHITECTURE

Fig. 5: System Architecture

In terms of creativity and difficulty, system and design are at the top of the life cycle. Together, the phrase "design" and "development process" characterise a completed system. These are the detailed technical requirements for putting the proposed system into action. A plan might be portrayed as "the most common way of applying different methods and standards to characterize a gadget, a cycle or a framework in adequate subtleties to allow its actual acknowledgment." Producing the output and presenting examples of both the input and result in the desired manner are the primary aims of the design. To fulfil the specifications of the intended output, it is necessary to construct the second set of input data and database files. Project development and testing take care of the processing step. As a last stage before deployment, management documents and evaluates information pertaining to the

system's rationale and an estimate of the prospective system's effect on users and the group. Just one word—"quality"—can describe how crucial software design is. In order to evaluate the quality of software, we need a representation, and design gives us just that. We run the danger of creating an unstable system that could fail if we don't design it first so that we can correctly transform client needs into a final software product or system. little tweaks are made, or ones whose quality isn't testable, or both. Development of a software product therefore includes this crucial step

6. IMPLEMENTATION

Service Provider

The Service Provider must provide their valid login credentials in order to access this module. Certain procedures will be available to him when he successfully logs in, including: Training and Test Data Sets, Browse All Remote Users, View Trained and Tested Accuracy in Bar Chart, View Results of Trained and Tested Accuracy, See Details of Predicted Crime Types, Locate Crime Type Ratio on Data Sets, Download Trained Data Sets, View Results of Crime Type Ratio.

Remote User

Number of users is present in this module. Registration is required prior to performing any operations. Details will be entered into the database after a user registers. Upon successful registration, he will be prompted to provide his authorised user name and password. Upon successful login, users will be able to do activities such as viewing their profile, posting crime data sets, and predicting the kind of crime.

7. ALGORITHMS

Support Vector Machine

One of the most notable regulated learning procedures, Backing Vector Machine (SVM) is utilized for both arrangement and relapseissues. Nevertheless, its main use is in Machine Learning classification tasks. Support vector machines (SVMs) plan to find the ideal choice limit or line that can separate n-layered space into classes, making it simple to allocate new information focuses to the right classification in the future. Hyperplanes are the finest choice boundaries. Support vector machines (SVMs) choose the hyperplane's extreme points and vectors.

Random Forest

Of the many AI calculations that fall under the umbrella of managed learning, Irregular Woodsstands out. This ML tool is versatile

enough to tackle classification and regression issues with ease. It relies on ensemble learning, which involves combining a few classifiers to fix a convoluted issue and lift the model's exhibition. As per the name, "Irregular Timberland is a classifier that contains various choice trees on different subsets of the given dataset and takes the normal to work on the prescient exactness of that dataset." The irregular woods doesn't rely upon a solitary choice tree yet rather midpoints out the gauges from all ofthem and then uses the one with the most votes to determine the final conclusion.

Decision Tree

Although Choice Tree is most frequently utilized to settle order issues, it is a directed learning approach that might be utilized to relapse issues too. In this sort of classifier, the qualities of a dataset are addressed by the inward hubs, the choice guidelines are addressed by the branches, and the outcome is addressed by each leaf hub. Two hubs, the Choice Hub and the Leaf Hub, go with up a choice tree. Leaf hubs are the consequences of decisions and have no more branches, as opposed to choice hubs, which might pursue any choice and have many branches. Highlights of the gave dataset are utilized to lead the test or make the decisions.

Naïve Bayes

An technique for tackling classification issues, the Credulous Bayes calculation is a sort of managed learning calculation that depends on Bayes hypothesis. Text classification utilizing a high-layered preparing dataset is its essential use. With regards to developing fast AI models that can create speedy expectations, one of the best and simple characterization calculations is the Gullible Bayes Classifier. It makes forecasts in light of the probability of an article'soccurrence; this makes it a probabilistic classifier.
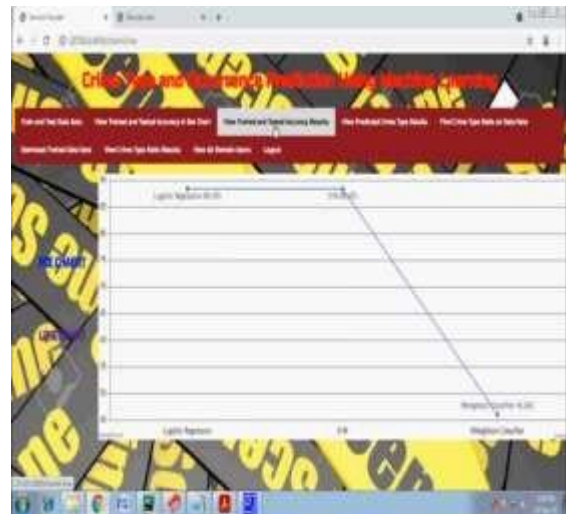
8. EXPERMENTAL RESULTS
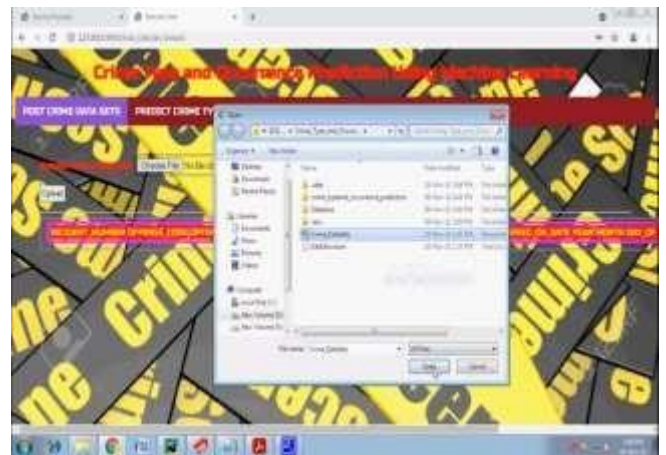
For analysis uploading date

8.4.Accuracy graph

after upload dataset display values of
dataset

dataset accuracy

find the accuracy for dataset

Accuracy graph for dataset
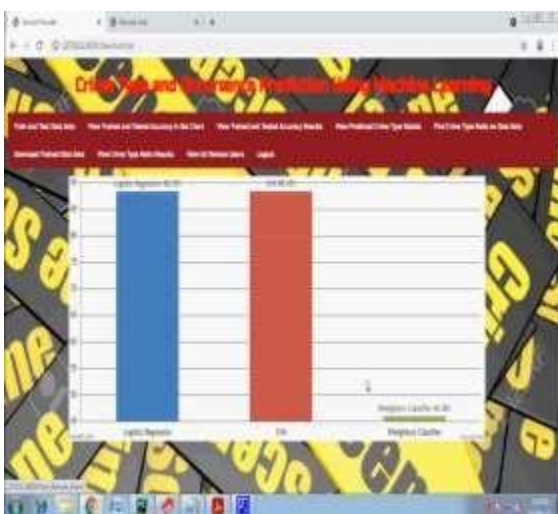
type of crimes display

type of crime graphs

9. CONCLUSIONS

Two classifiers, such as Multi nominal NB and Gaussian NB, help to overcome the issue of dealing with real valued characteristics and nominal distributions. It is ideal for making predictions in real time and doesn't need a lot of preparing time. Moreover, it addresses the issue of managing consistent objective arrangements of factors, which waspreviously an insurmountable obstacle in the previous work. Therefore, Naïve Bayesian Arrangement may be utilized to estimate and recognize the most well-known violations. Using certain conventional measurements, we can also calculate the algorithm's performance. When evaluating algorithms, the most important parameters to consider are accuracy, F1 score, recall, and average precision. Applying machine learning

methods would significantly improve the accuracy value.

10. Future Enhancement

It has several limits, but it does solve the issue with the previous work. The estimation's probability is 0 in the case when class labels are not present. The use of additional machine learning categorization models has shown to improve overall performance and accuracy in crime prediction, suggesting that this area of study might be further upon in the future. Incorporating neighbourhood income data helps with future improvement studies by allowing researchers to predict if there is a correlation between a person's income and the crime rate in their area.

11. REFERENCES

[1] Suhong Kim, Param Joshi, Parminder Singh Kalsi,Pooya Taheri, "Crime Analysis Through Machine Learning", IEEE Transactions on November 2018.

[2] Benjamin Fredrick David. H and A. Suruliandi,"Survey on Crime Analysis andPrediction using Data mining techniques", ICTACT Journal on Soft Computing on April 2012.

[3] Shruti S.Gosavi and Shraddha S. Kavathekar,"A Survey on Crime Occurrence Detection and prediction

Techniques", International Journal of Management, Technology And Engineering , Volume 8, Issue XII, December 2018.

[4] Chandy, Abraham, "Smart resource usage prediction using cloud computing for massive data processing systems" Journal of Information Technology 1, no. 02 (2019): 108-118.

[5] Learning Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimparia, "Crime Pattern Detection, Analysis & Prediction using Machine", International Research Journal of Engineering and Technology, (IRJET) e-ISSN: 2395-0056, Volume: 07, Issue: 06, June 2020

[6] Umair Muneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir and Hafiz Husnain Raza Sherazi, "Spatio-Temporal Crime Hotspot Detection and Prediction: A Systematic Literature Review", IEEE transactions on September 2020.

[7] Nasiri, Zakikhani, Kimiya and Tarek Zayed, "A failure prediction model for corrosion in gas transmission pipelines", Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, (2020).

[8] Nikhil Dubey and Setu K. Chaturvedi, "A Survey Paper on Crime Prediction Technique Using Data Mining", Corpus ID: 7997627, Published on 2014.

[9] Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abdi and Abdulrahman Al-Ahmari, "Computational System to Classify Cyber Crime Offenses using Machine Learning", Sustainability Journals, Volume 12, Issue 10, Published on May 2020.

[10] Hyeon-Woo Kang and Hang-Bong Kang, "Prediction of crime occurrence from multimodal data using deep learning", Peer reviewed journal, published on April 2017.