ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-08
Dr.MD.Rafeeq *et. al.*, / International Journal of Engineering & Science Research

Crime Type and Occurrence Prediction Using Machine Learning

Dr.MD.Rafeeq[1], M Vivek[2] Adicherla Prathiba[2], Mamidi Yuvaraj[2], Dolharkar Maithlee Narendra [21]*Professor,Computer Science and Engineering, CMR Engineering College, medchal, T.S, India*[2]*B.Tech, Computer Science and Engineering, CMR Engineering College, medchal, T.S, India*

Abstract: Crime has emerged as a clear means of putting individuals and society in difficulty in our recent period. An increase in crime causes an imbalance in a nation's population. Understanding crime trends is essential for analysing and anticipating this kind of criminal activity. Using open source data from Kaggle, this research applies one such crime pattern analysis, which is then utilised to forecast the most recent incidents. This project's main goal is to determine which kind of crime contributes the most, along with when and where it occurred. The accuracy gained was quite good when compared to pre-written works, and several machine learning methods, such as Naive Bayes, are used in this study to categorise among different criminal patterns.

1. INTRODUCTION

Crime has emerged as a significant issue that is thought to be intensifying quickly. When an activity is against the law, is very offensive, and breaches a rule, it is considered to be a crime. The knowledge of criminology's many facets as well as the ability to identify patterns are both necessary for the crime pattern analysis. Incorporating technology to control some of these illicit activities requires a lot of time and effort on the part of the government. Therefore, it is necessary to employ machine learning methods and its records in order to forecast the kind and patterns of crime. It mandates the use of pre-existing crime data and makes predictions about the kind of crime that will occur and when based on the place and time. Numerous studies were conducted by researchers to aid in the analysis of crime trends and their relationships in a particular area. Some of the hotspots that were examined have made it simpler to categorise the crime trends. This makes it possible to help the authorities address them more quickly. This method makes use of a dataset from Kaggle open source that is based on numerous parameters as well as the time and place where they occur during a predetermined period of time. We suggested a categorization method that aids in identifying the types of crimes and concentrations of criminal activity that occur at certain times of day. A machine learning approach is suggested to be used in this case to match criminal patterns with their category using the provided temporal and geographical data.

2. LITERATURE SURVEY

2.1 Existing System

Pre-work is the process of removing duplicate values and characteristics from a dataset that was collected from an open source. Decision trees have been used to extract characteristics from massive amounts of data and to identify patterns in crime.[1] [2] It offers a basic framework for subsequent categorization procedures. Deep neural networks are used to

extract features from the identified crime patterns. The performance is computed for both training and test values based on the prediction [3] [4]. The crime prediction aids in foreseeing future instances of any kind of criminal activity and assists the authorities in swiftly resolving them.

2.2Proposed System

The data is initially pre-processed using the filter and wrapper machine learning techniques in order to eliminate redundant and unnecessary data values. Additionally, it lowers the dimensionality; as a result, the data is now clean. After then, the data passes through another dividing step. It is divided into a training data set and a test data set. Both the training and testing datasets are used to train the model.The next step is mapping.[5] [6] To make categorization simpler, the crime type, year, month, time, date, and location are all mapped to integers.

- Utilising Nave Bayes, the independent relationship between the qualities is first investigated. The retrieved independent characteristics are classified using Bernouille Nave Bayes. It is possible to assess the incidence of crime at a certain time and place by labelling the criminal aspects. The most common crimes are finally discovered, together with geographical and temporal data. The accuracy rate calculation is used to evaluate how well the prediction model is working. Python was used to create the prediction model, which was then executed on Colab, an online compiler for machine learning and data analysis models.

2.2 Pre-processing of Data

To avoid pointless infractions, pre-processing of data collected from the open source is important. The dataset was selected for the city of Denver because it has copious amounts of crime data over six years. The missing integral in the supplied attribute values is indicated to be found using the machine learning approach filter and wrapper. Data cleansing is essential for both the performance of the started process and for training a prediction model.[7] [8]

Instances are filtered, and unnecessary context is removed from datasets. The filtering techniques aid in determining the importance of the characteristics. When choosing a feature, the association with the dependent values is taken into account. By actually training a prediction model on the feature subset, the wrapper technique is used to assess how helpful it is. Preprocessing separates the data into training and test characteristics.

Fig 1: System Flow

3. SYSTEM ANALYSIS&DESIGN

3.1Mapping

The characteristics of the crime, such as the kind of crime, the day and time the crime happened, are first separated. It is then translated to an integer for straightforward labelling. The tagged information is subsequently examined and utilised for graph plotting. Python is used as the programming language to accomplish the suggested task since it is ideal for machine learning. To produce a graph to display the frequency of criminal activity, the software matplotlib is loaded. The most often occurring crime may be plotted on the graph, which helps with future prediction.[9] [10]

3.2 Neural Network Classification

The use of Naive Bayes is justified since geographical and temporal data are often included in crime prediction. Since the chosen crime characteristics have an independent influence on them, the independent effect among the attribute values is initially examined. They are trained using criminal data relating to robbery, burglary, homicide, sexual abuse, armed robbery, chain snatching, gang rape, and highway robbery before being employed in the creation of a model. There have been several indicated expanded Naive Bayes approaches.[11]

The Gaussian The selection of real valued attributes is connected to Naive Bayes. It is also known as a normal distribution, and it is done by determining the mean and standard deviation from training data.

2. For numerous classifiers that correspond to the categorical features in the learned value, multi-nominal Nave Bayes is employed.

Bernouille 3. For the independent feature impacts of the chosen characteristics for crime prediction, Naive Bayes is applied.

ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-08
Dr.MD.Rafeeq *et. al.*, / International Journal of Engineering & Science Research

D. Prediction of Crime

By expanding the supported criminal attributes, the anticipated crime type is projected. The nominal values are then affected by the characteristics. A single tuple might be used as an example to effectively illustrate it.

Thinking about a tuple

1. "Gateway Town, October 20, 2020, 2: 30 PM, Friday" translates to "Larceny" as a crime that entails stealing someone else's property.

Based on the characteristic retrieved, take into account the likelihood of occurrence:

1. "Gateway town" = "Theft has taken place"
2. "October" => "There has been a theft"
3. {2020} => There has been a theft.
4. "2:30 PM" => "There has been a theft"
5. "Friday" = "There has been a theft"

The conditional probability is computed once the independent occurrence has been created. So, we could anticipate the sort of crime.

the use of symbols

1. m stands for Month. 2. t stands for Time.
3. a denotes Area.
4. D stands for Day.
5. y symbolises the year, while 6. c stands for kind.

INCIDENT_ID	OFFENSE_ID	OFFENSE_CODE	OFFENSE_CODE_EXTENSION	OFFENSE_CATEGORY_ID
[illegible]	[illegible]	2399	0	theft-other
[illegible]	[illegible]	5707	0	criminal-trespassing
[illegible]	[illegible]	2399	1	theft-bicycle
[illegible]	[illegible]	2308	0	theft-from-bldg
[illegible]	[illegible]	5016	0	violation-of-restraining-order

Table 1. Dataset Collection

FIRST_OCCURRENCE_DATE	LAST_OCCURRENCE_DATE	REPORTED_DATE
12/27/2018 5:58:00 PM	NIL	12/27/2018 6:51:00 PM
01-06-2021 9:20:00 PM	NIL	01-07-2021 12:23:00 AM
06-08-2017 1:15:00 PM	06-08-2017 5:15:00 PM	06-12-2017 8:44:00 AM
12-07-2019 1:07:00 PM	12-07-2019 6:30:00 PM	12-09-2019 1:35:00 PM
12/22/2018 6:15:00 PM	12/22/2018 8:31:00 PM	12/22/2018 10:00:00 PM

Table 2. Crime Dataset with occurrence date and time

NEIGHBORHOOD_ID	IS_CRIME	IS_TRAFFIC
montbello	1	0
Gateway-green-valley-ranch	1	0
wellshire	1	0
belcaro	1	0
cherry-creek	1	0

Table 3. Neighbourhood dataset

3.3 System Architecture: The performance of the implied prediction is then evaluated in order to achieve a high degree of accuracy when compared to the pre-existing model used. The training is done with cross validation that helps in training the data on different set of training data. It will evaluate the accuracy for overall splits in the cross validation implied. In python, in order to calculate the value of accuracy we need to pass the data arguments such as model name, target set and cv that helps in signifying the split occurrence. Finally, the mean and the standard deviation of the average precision is calculated. The accuracy of 93.07% has been achieved who gives a great increase in compared to existing prediction models.

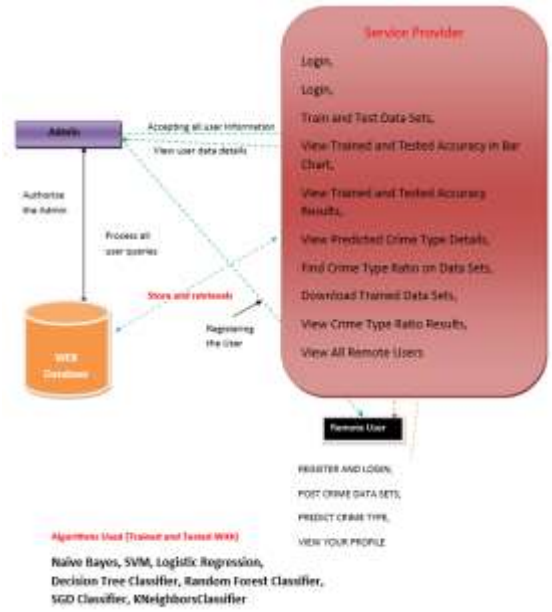

Fig 2 : System Architecture

3.4 Data Flow Diagram : Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

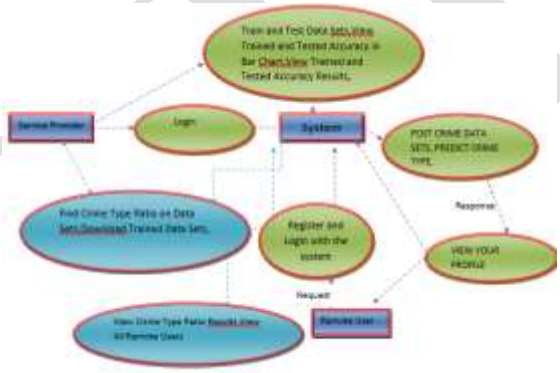

Fig 3: Data Flow Diagram

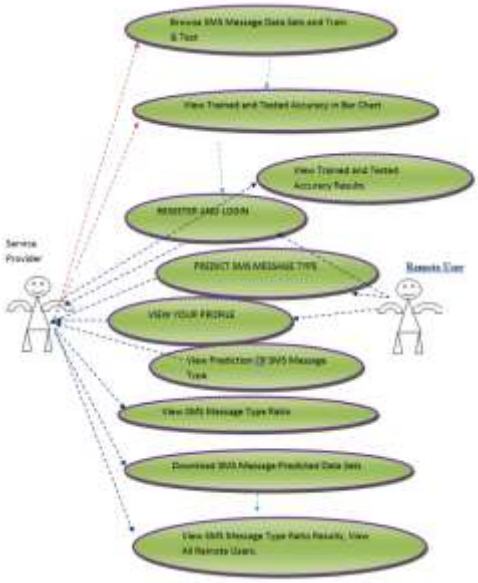

Fig 4 Use Case UML Diagrams

5. CONCLUSION

In this study, two classifiers—Multinominal NB and Gaussian NB—are used to address the challenge of dealing with the nominal distribution and real valued features. Real-time forecasts can be made with the least amount of training time possible. Additionally, it solves the issue of dealing with continuous target sets of variables, which the earlier research refused to fit. Therefore, Nave Bayesian Classification might be used to forecast and identify the crimes that occur the most often. Additionally, the algorithm's performance is determined using a few common criteria. The metrics average precision, recall, F1 score, and accuracy are the ones that matter most when evaluating an algorithm. By using machine learning methods, the accuracy value might be improved significantly.

6. FUTURE SCOPE

It has certain shortcomings even if it fixes the issue with the previous effort. If there are no class labels, then there is no chance that an estimate will be made. The use of additional machine learning categorization models as a future expansion of the suggested study will improve performance overall and raise the accuracy of crime prediction. By considering the income data for neighbourhood places, it aids in the provision of a better study for the future improvement by allowing one to predict if there is a correlation between the income levels of a certain neighbourhood place and its crime rate.

7. 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, MuzamilKacchi, PranaliGavali 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, FadratulHafinaz Hassan, Mubashir Ali, AneesBaqir 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", Peerreviewed journal, published on April 2017

[11] Shantala S Yalawar K Vijaya Babu, Mrutyunjaya S Yalawar, A New Approach for Secure Login Method and Forestall Cyber bulling in Social Media, 2019/5, International Journal of Recent Technology and Engineering (IJRTE),Volume 8, Issue Ic2,1243-1246. https://scholar.google.co.kr/citations?view_op=view_citation&hl=en&user=Tyqm5YUAAAAJ&citation_for_view=Tyqm5YUAAAAJ:UebtZRa9Y70C