Machine Learning based Algorithms for Predicting the Criminal Activity Nature and its Frequency

Mrs. Nallanichakravarthula Ananthasrivyshnavi[1]
Mrs.Y.Susheela[2], Dr.N.Chandra Mouli[3]
[1]M.Tech Student
[2]Associate Professor, CSE Dept.
[3]Professor and HOD CSE Dept.
Vaageswari College of Engineering
Karimnagar, Telangana Sate, INDIA

Abstract:

Using government benefits without remitting payment has been more common in recent years. This lifestyle is dangerous. If criminal activity increases across the country, there may be social disruption. To effectively dissuade criminal activity, it is necessary to understand the basic techniques used by criminals. This project uses criminal pattern analysis on publicly available crime data from Kaggle. The acquired data is then used to produce accurate predictions about criminals' future behaviour. The study's goal is to determine the specific places and times of day when criminal behaviour is most frequent. In this study, Naive Bayes machine learning was used to categorize criminal activity. According to a prior study, its accuracy is lower than that of the categorization model.

Keywords: Machine Learning, Service Level Agreements, enhanced decision tree algorithms

1.Introduction

Considering the prevalence of criminal behaviour in today's society which is occurring often an utmost importance has arise that every facet of criminology to be investigated with comprehensive details. In order to make accurate predictions about criminal behaviour, it is necessary to have access to a large number of datasets and to employ newly developed advanced machine learning based techniques. This method of analysing past criminal records that can be helpful in estimating the frequency of future criminal episodes as well as the occurrence locations of those crimes. The results of a great number of scientific investigations have unequivocally demonstrated a connection between illegal activities that take place in various regions of the world. The consequence of this is that it is now possible to analyse and comprehend this phenomenon. Websites from a wide variety of categories have been shown to be important sources of information and findings regarding patterns of criminal behaviour. The result of this is that law enforcement will be able to resolve cases in a more expedient manner. This investigation makes use of the Kaggle database, which is accessible to the general public. As part of the compilation process, the times, locations, and dates of all of the events that took place during a particular time period were compiled into a single document. The findings from this research offer a methodology that may be utilized to identify and evaluate vulnerable areas that have consistently high crime rates. Through the utilization of time and location data, machine learning algorithms have successfully identified actions that are consistent with one another.

2. Literature Review

Suhong Kim, Param Joshi, Parminder SinghKalsi, Pooya Taheri (2018), determine the level of precision that Machine Learning (ML) based algorithms may achieve in detecting instances of maladaptive behaviour. The primary goal of this inquiry is to determine the extent of criminal activity occurring in metropolitan cities. Machine learning based techniques such as enhanced decision tree algorithms(EDTA) and K-nearest neighbour algorithms(KNA) were used to forecast the crime rate that is going to occur, When applied it reveal the future estimates, these models have an accuracy

range of 39% to 44%. Benjamin Fredrick David. H and A.Suruliandi (2017), analysed Police forces around the world routinely use statistical analysis to predict future crime occurrence. An examination of historical and contemporary trends in unlawful behaviour using publicly available data. This methodology focuses on analysing historical data using computer based algorithm tools, statistical models, and deep learning techniques. The projected level of accuracy and precision for each method is taken into account while rating and grouping the numerous ways used in different disciplines in this study. Shruti S.Gosavi and Shraddha S.Kavathekar (3), it is critical to keep detailed entire records and be prepared ready to respond quickly without any time lapse. Deep neural networks, association rule mining, naive Bayes, decision trees, and artificial neural networks are some of the numerous interconnected computer models and methodologies. The majority of offenses take place in specific places. Post-processing algorithms can be used to predict where unlawful conduct is most likely to occur. Chandy and Abraham (4), the study looked into the resource use trends of a cloud computing-enabled firm. There were no issues as a result of either excessive or insufficient resource consumption. Excessive use of resources produces more waste, raising expenses for those affected. Service Level Agreements (SLAs) are being breached due to insufficient supply, indicating a degradation in the quality of service offered.

2. System Design Feasibility
A feasibility study determines the likelihood that a given endeavor will be completed successfully. An assessment will be done to determine the feasibility and practicability of the proposed strategy. Implementing the suggested resolution will result in long-term cost savings for the group.

2.1 Economical Feasibility
The analysis calculates the level of financial volatility that the company will experience as a result of the technology installation. Because of financial constraints, the firm is unable to devote significant resources to research and development efforts targeted at producing unique products.

2.2 Technical Feasibility
The primary goal of it is to have a thorough assessment and evaluation of the chosen research topic. All stakeholders gain from avoiding excessive network traffic during the implementation of a new system. Slight change to the existing process is required to properly address the current difficulty.

2.3 Proposed System
Before removing superfluous information, data is sorted and classed using learning algorithms. A data loss has happened as a result of the decrease in traffic. The information is then divided into different categories. Create a diverse set of samples for evaluation and education purposes. During training and evaluation, the model is exposed to various data kinds. Every event is allocated a unique number value. In addition to the yearly, monthly, diurnal, and hour components, this indicator includes information about the nature of the event.

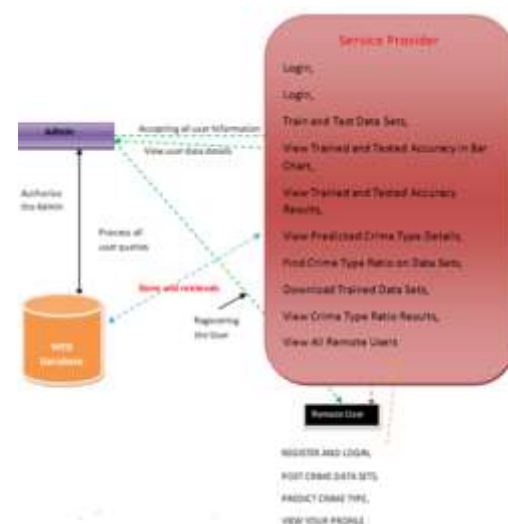## 3.System Design and Implementation

Figure 1. System Architecture

The Python interpreter executes code entered by the user or opened from a Python program file. It is possible to manage tasks, modify code, and correct errors using the various tools available in integrated development environments (IDEs). Fig.1 each components

serves a variety of purposes. These libraries facilitate routine computer tasks by containing prewritten code. The abundance of tools and information available in the modern era allows us to deal with a wide range of situations and challenges. Changes to user-generated information are not required for academic compliance. Matplotlib is widely acknowledged to be capable of visualizing and simplifying complex data. It can produce spreadsheets, bar graphs, histograms, and a wide range of other images and graphs. Matplotlib allows you to change the composition, color, and label of your diagram. As a result, creating visually appealing data visualizations becomes much easier.

The Soup program is frequently used to modify previously downloaded data from the internet. Training and testing data collection is the first stage in the machine learning process. Scikit-learn, pandas, NumPy, and other Python utilities make it simple to perform other basic tasks like data collection and cleaning. files, computer APIs, and other APIs are just a few of the places data can be gathered from. Also, additional APIs can be used as sources of information. For testing and training purposes, the software can divide a dataset into multiple groups and handle missing data values and feature proportions.

4. System Testing

Testing is primarily concerned with identifying and validating any potential issues that may arise. Multiple iterations of testing help to detect and resolve bugs and other issues that may arise in a product. This approach is adaptable and can be used in a variety of sectors, including collections, individual components, and entire products. To gain a thorough understanding of how software works, it must be subjected to rigorous testing in accordance with scientific methodology. The primary goal of software testing should be to ensure that the final iteration performs as expected.

The application meets its objectives while exceeding the expectations of its users. At this time, no significant developments have occurred that would diminish the importance of the current point of contention.

4.1 Testing Strategy

It's possible for software to work better if it follows strict design principles and makes a lot of thorough system test cases. This makes it easier to come up with a good testing strategy for the system. To make sure your testing strategy works, the steps of creating test cases, running tests, and analysing results must all work together without any problems. There are many different testing methods that need to be used to fully evaluate software. These methods range from the simplest to the most complex. A small amount of testing is done to make sure the code is in the right place. Important parts of the system are carefully checked to make sure they work right.

The software must be put through a lot of tests to make sure it meets high standards of quality. People are better able to understand and get information about important laws and rules now that there are more media platforms available. As expected, there is a lot of interest in the subject because it is so important. It is important to carefully look over the proposed system before starting user acceptance testing.

5. Results

The out are depicted in the figures from 2 to 4 using machine learning based algorithm for predicting the criminal activities and also its frequency. The regression analysis is shown in Fig.2 and algorithm prediction analysis is given in Fig.3 and crime type with good precision of occurrence is shown in Fig.4
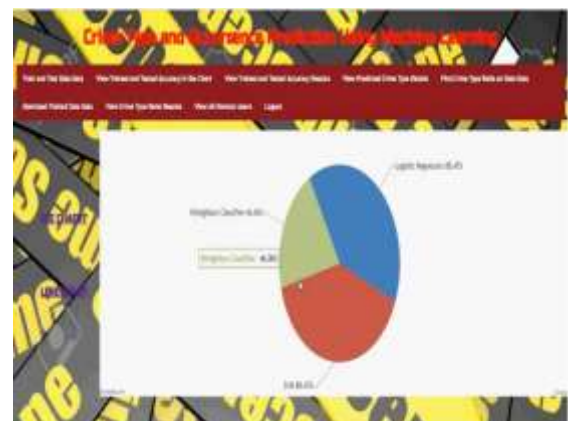
24   Vol.19, Issue. 2, July-December : 2024

Figure 2. Trained and Tested Accuracy in Pie Chart

Figure 3. Crime Type Prediction Details

Figure 4. Crime Type Details in Line Chart

6. Conclusions

The Multi-Nominal Naive Bayes (NB) and Gaussian Naive Bayes (NB) models are used to look into the problem of nominal distribution vs. real-valued characteristics. These kinds of predictors are very good at figuring out what will happen. The suggested answer takes into account the objective variables' continuous properties in order to solve a problem that has already been found. A lot of common bad behavior can be found and stopped with Naive Bayesian Classification. One common way to figure out how well an algorithm works is to compare its results to established norms in the same field. A lot of the time, average precision, good score, and accuracy are used to judge how well a machine learning is functioning.

References

[1] Suhong Kim, Param Joshi, Parminder SinghKalsi and Pooya Taheri, "Crime Analysis Through Machine Learning," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2018, pp. 415-420, doi: 10.1109/IEMCON.2018.8614828.

[2] H. Benjamin Fredrick David1 and A. Suruliandi, Survey on Crime Analysis and Prediction Using Data Mining Techniques, ICTACT Journal on Soft Computing, April 2017, Volume: 07, Issue: 03, ISSN: 2229-6959, ISSN: 2229-6959 (ONLINE) Pp:1459 - 1466

[3] Shruti S Gosavi and Shraddha S.Kavathekar A Survey on Crime Occurrence Detection and prediction Techniques, International Journal of Management, Technology and Engineering , Vol.8, Issue XII, Dec.2018, ISSN : 2249-7455, pp:1405-1409.

[4] Abraham Chandy, Smart Resource Usage Prediction Using Cloud Computing For Massive Data Processing Systems, December 2019, Journal of Information Technology and Digital World 01(02):pp:108-118

[5] Rohit Patil, Muzamil Kacchi, Pranali Gavali, Komal Pimparia, Crime Pattern Detection, Analysis & Prediction using Machine Learning, International Research Journal of Engineering and Technology, Vol.: 07 Issue: 06 | June 2020, ISSN: 2395-0072, pp:119-121

[6] Umair Muneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir Hafiz Husnain Raza Sherazi, Spatio-Temporal Crime HotSpot Detection and Prediction: A Systematic Literature Review, September 2020 , IEEE Access ,pp: 166553 – 166574, ISSN: 2169-3536

[7] Nikhil Dubey, Setu Kumar Chaturvedi, A Survey Paper on Crime Prediction Technique Using Data Mining , Int. Journal of Engineering Research and Applications, ISSN : 2248-9622, Vol. 4, Issue 3(Version 1), March 2014, pp.396-400

[8] Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abidi and Abdulrahman Al-Ahmari Computational System to Classify Cyber Crime Offenses using Machine Learning, Sustainability 2020, 12(10), 4087, pp"2-16. ISSN: 2071-1050

[9] Hyeon-Woo Kang , Hang-Bong Kang, PLOSONE Prediction of crime occurrence from multi-modal data using deep learning April 24, 2017, vol.12, issue:4,pp:1-19.

[10] Prabakaran, S.; Mitra, S. Survey of Analysis of Crime Detection Techniques Using Data Mining and Machine Learning. J. Phys. Conf. Ser. 2018, vol.10, issue:12, pp.10-17.