ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

Computational Procedure For Predicting Crime Data Using Machine Learning Models

[1]Dr. N. Shanmuga Priya, [2]Mr. S. Vikass, [3]Mr. VF. Victor Noah

[1]Associate Professor and Head, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[2]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

[3]PG Student, II MCA, Department of Computer Applications, Dr. SNS Rajalakshmi College of Arts & Science, Coimbatore.

Abstract: -

A crime is a deliberate act that can cause physical or psychological harm, as well as property damage or loss, and can lead to punishment by a state or other authority according to the severity of the crime. The number and forms of criminal activities are increasing at an alarming rate, forcing agencies to develop efficient methods to take preventive measures. In the current scenario of rapidly increasing crime, traditional crime-solving techniques are unable to deliver results, being slow paced and less efficient. Thus, if we can come up with ways to predict crime, in detail, before it occurs, or come up with a "machine" that can assist police officers, it would lift the burden of police and help in preventing crimes. In this paper, we describe the results of certain cases where such approaches were used, and which motivated us to pursue further research in this field. The main reason for the change in crime detection and prevention lies in the before and after statistical observations of the authorities using such techniques. The study provides access to the datasets used for crime prediction by researchers and analyzes prominent approaches applied in machine learning algorithms to predict crime, offering insights into different trends and factors related to criminal activities. The main reason for the change in crime detection and prevention lies in the before and after statistical observations of the authorities using such techniques. By gaining a deeper understanding of crime prediction techniques, law enforcement agencies can develop strategies to prevent and respond to criminal activities more effectively.

Keywords: *Crime Prediction, Crime Detection, Crime Datasets, Machine Learning, Crime Analysis.*

1. INTRODUCTION

A crime is a form of violence or illegal act done by a perpetrator against another person that can cause harm or property damage and is punishable by the law of the governing state of authority in which the crime was carried out. Law authorities apply crime-solving techniques to take preventive measures but in many cases, they cannot deliver effective results [1]. Data mining and ML are both versatile fields that involve the use of computers and mathematics where the programming is completed for the system to perform certain tasks, these are both important parts of crime prevention and detection Data mining can be considered as the process where discovers of new patterns from large data sets involving methods from statistics and AI, but also database management. The availability of enormous volume of data being made available by certain governments has given motivation to researchers to further pursue research in the field of crime. Historical data has made it an interesting subject that sparked attention in research; many researchers have proposed several different models for predicting the future occurrence of crimes [2]. In some areas law authorities have restrictions over their data and may not make this available to researchers in the area, causing further frustration and disappointment.

There are many different types of crimes ranging from antitrust offenses to environmental violations, frauds and beyond. Crime imposes enormous financial, physical and social harms on individuals, communities and society in general. Because of their special characteristics and techniques by which they are committed, they pose significant problems for law enforcement and regulatory agencies interested in controlling them. Evidence suggests that crime is pervasive, Widespread and growing [3]. In order to control the cyber crime activities, the law enforcers have to effectively meet out some challenges of crime control and maintenance of public order. One is those take the network as criminal objects and the other one is using those network to commit crime such as terrorism, fraud, and illegal trade etc., therefore it is essential to create database for crime and criminals. Data mining helps in exploring those large database and makes it convenient for the user and organization. Nowadays the world is facing the problems such as trafficking, smuggling, kidnapping and terrorism. To prevent these miner analyze the web information from the perspective of events and apply some research results related to the events to solve the problem of web crime mining.

Machine learning (ML) is a subfield of Artificial Intelligence being used across many different fields today to predict the future occurrence of certain events as well as better decision making [4]. ML can be understood as the study of computer algorithms that can automatically improve on their own through experience/learning and by the use of data. Deep Learning (DL) is a subset of machine learning that is inspired by how our brains function, this technique is an artificial neural network that includes many different layers and layer types (e.g., pooling layer, convolution layer, fully-connected layer, dropout layer) that attempt to replicate the behavior of our brains. There exist four types of learning types, which are supervised, semi-supervised, unsupervised, and reinforcement learning. AI comprises both computer and mathematics (i.e., statistics) aspects where the programming is performed for the system to perform a certain action, commonly associated with humans [5].

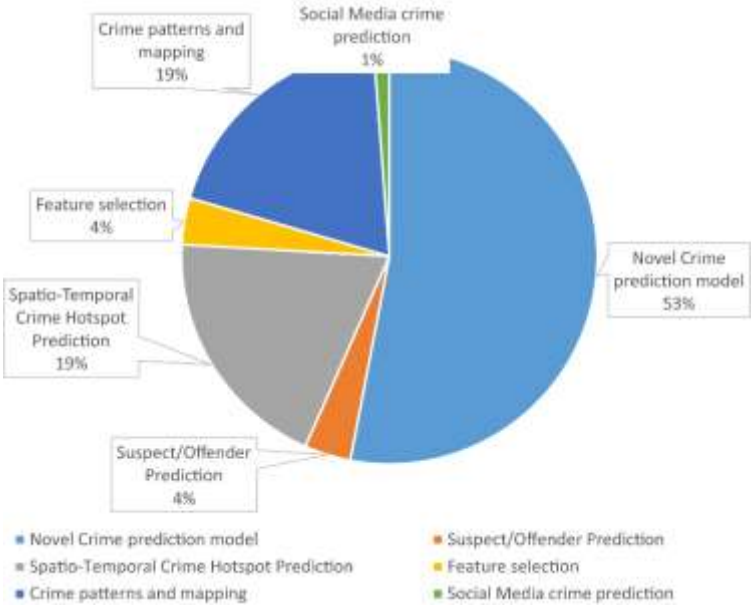

Fig.1.1: - Crime Data Analysis

Data mining is process of applying various methods such as clustering, decision tree etc., to data with the intention of uncovering hidden patterns. A primary reason for using data mining is to assist in the analysis of collections of observations of behavior. Data mining proposes several classification techniques which can be effectively applied to detect

fraudulent transactions. The main objective of this is to develop a security application for detection of cybercrime based on proposed system in two stages. First Stage is cybercrime detection systems, which collects and analyze the data and second stage is to reduce the false alarm rate [6].

1.1. Prevention of Crime

Cyber crimes can occur locally within a country and globally across borders. In India, the statistics of the incidents of cybercrimes are documented state wise and on a yearly basis. Thus it is easy to make a state wise analysis of cybercrimes rates. The police forces available in various regions differ from state to state. Thus state wise analysis is also effective because the results of the analysis can be used to motivate the police force towards mitigating the particular type of cybercrimes prominent for that state [7]. The governments usually establish organizations such as courts; prosecutions and police are responsible for the maintenance of law and order. These organizations are responsible to curb the rate and occurrence of crimes. The crime prevention methods are

- Safeguarding the life and property of the society.
- Crime elimination rules out litigation, which follows along the way of sensing a crime.
- Prevention also saves the authorities from the difficulties without doing crime all the time and using immediate activity for the investigation.

2. RESEARCH METHODOLOGY

Systematic literature reviews (SLRs) are and have been used in many different fields of academia from engineering to medicine to gather and summarize data on a certain research topic, we employed guidelines to follow a structured approach for our study [8]. By using an SLR, we can also identify the challenges and possible solutions that can be employed. It defines the guidelines for the following three (3) main phases:

- Planning the review: Gathering related data and research work related to our research topic (i.e., the use of machine learning in crime prediction), defining research questions and systematic search protocols that include the selection of keyword strings to be used for related papers, and how this criterion will be applied to the papers. The planning phase also includes the following phases:
  - Identification of the need for a review.
  - Commissioning a review.
  - Specifying the research question(s).
  - Developing a review protocol.
  - Evaluating the review protocol.
- Conducting the review: Putting a classification schema in place, details on how the data and papers will be separated for analysis where features are grouped based on similar or common attributes. The papers are subject to inclusion and exclusion criteria. The papers that pass the criteria should meet the minimum quality assessment threshold, which is mostly selected as the mean value. Critical data from these papers are extracted and synthesized to present a general overview of the understanding of how machine learning can be used for crime prediction to identify possible gaps and opportunities in the selected field. The phases in the conducting phase include:
  - Identification of research.
  - Selection of primary studies.
  - Study quality assessment.
  - Data extraction and monitoring.
  - Data synthesis.

- Reporting the review: The final step discusses and presents our findings, the research questions which were identified in step 1 are addressed and visually presented with graphs, figures, and tables ß if needed and may include:
  - Specifying dissemination mechanisms
  - Formatting the main report
  - Evaluating the report

The challenge in data mining crime data often comes from the free text field. While free text fields can give the newspaper columnist, a great story line, converting them into data mining attributes is not always an easy job [9]. Here will look at how to arrive at the significant attributes for the data mining models.

- **Data mining crime:** Data mining deals with the discovery of unexpected patterns and new rules that are "hidden" in large databases. Data mining is used to improve crime analysis and aid in reducing and preventing crime. It is the powerful software in the study of crime cases, which mainly concerned entity extraction, clustering, classification, and network analysis. By doing the continues crime acts the data of the criminals are fetched up frequently.
- **Object extraction:** In object extraction the attributes and relationship of those attributes are extracted. The main objective of this methodto fetches the information from the database. This method append on the concept of the data mining techniques. The identity of the person, address, cases and some personal issues about the person has been automatically done by the entity extraction.
- **Focus on data:** The main focus of this analysis is on web mining of content with the help of clustering techniques. Clustering will convert nonlinear statistical relationship between high dimensional data into simple geometrical relationship in low dimensional display.

2.1. Classification of Crime

Data mining techniques are used to predict the patterns in the incidents of various types of cybercrimes. Feature selection techniques are used to determine the prominent types of cyber crimes [10].

CRIME: Crime is "an act or the commission of an act that is forbidden or the omission of a duty that is commanded by a public law and that makes the offender liable to punishment by that law".

The internet as tools can classify the crime as follows

- Traffic Violations: - Driving under the influence of alcohol, fatal / personal injury / property damage traffic accident, road rage.
- Sex Crime: - Sexual offences.
- Fraud: - Forgery and counterfeiting, frauds, embezzlement, identity deception.
- Arson: - Arson on buildings.
- Drug Offences: - Narcotic drug offences.
- Violent Crime: - Criminal Homicide, armed robbery, aggravated assault, other assaults.
- Cyber Crime: - Internet frauds, illegal trading, network intrusion / hacking, virus spreading, hates crimes, cyber piracy, cyber pornography, cyber-terrorism, theft of confidential information [11].

3. DATA MINING TECHNIQUES AND IMPLEMENTATION

Data mining techniques such as association analysis, classification, clustering analysis are used to identify patterns in structured data. The new use of data mining in this paper is to give the structured data from unstructured data. Entity extraction identifies particular patterns of data such as text, images and audio data etc. Entity extraction provides basic

information for crime analysis, but its performance depends greatly upon the availability of excessive amount of clean input data [12, 13].

Clustering: Clustering techniques group data items into classes with similar characteristics to maximize or minimize similarity. Clustering crime incidents can automate major part of crime analysis but it is limited by the high computational intensity typically required.

3.1. K-Means Clustering Algorithm

The K-means [14] algorithm is an evolutionary algorithm that gains its name from its method of operation. The algorithm clusters observations into K groups, where K is provided as an input parameter. It then assigns each observation to clusters based upon the observation proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again. The working of algorithm is explained as follows:

1. The algorithm arbitrarily selects K points as the initial cluster centers.
2. Each point in the data set is assigned to the closed cluster based on the Euclidean distance between each point and each cluster center.
3. Each cluster center is recomputed as the average of the points in that cluster.
4. Steps 2 and 3 repeats until the clusters converge. Convergence may be defined differently depending upon the implementation, but it normally means that either no observation change cluster when step 2 and 3 are repeated or that the changes do not make a material difference in the definition of the clusters [15].

3.2. Implementation:

Experiments are conducted on real world data of few users as well as on the synthetically generated data of various users with different kind of usage behavior [16]. The collected data is distributed into three categories, two third of genuine data GD and fraud history data FD are used for training and remaining datasets are used for prediction and web crime detection. The various implementation steps are:

Step1: Data collection

Corporate/Companies are not ready to share information of their employees. Therefore the performance of the proposed system has been tested on five real users' data and on synthetically generated data by the simulator. For describing the implementation steps we select three legal and two crime data sets which are given in table 3.1.

Table 3.1: Sample Data

Users	A1	A2	A3	A4	A5	A6
A	H	L	H	M	H	M
B	M	H	L	M	L	H
C	H	H	M	H	L	M
D	M	M	M	L	L	H
E	L	H	M	M	H	L

Step2: Select the Sample data

The details of the persons can be chosen appropriately. In the case of existing crime detection large profile details increases the accuracy but at the same time it increases the training time.

Step3: Data Cleaning

Select only the necessary data and neglect the unwanted data. Sample transaction data used for training after cleaning is given in table 3.1.

Step4: Data Transformation

Classify the data into two clusters- cluster 0(genuine user) and cluster I (illegal data detected) using K-means clustering algorithm. This work selects K as two. After executing K-means algorithm on the sample transaction in table I, we get the cluster which is given in Table 3.2.

Table 3.2: Result of Clustering

Users	A1	A2	A3	A4	A5	A6	Cluster
A	H	L	H	M	H	M	Cluster1
B	M	H	L	M	L	H	Cluster1
C	H	H	M	H	L	M	Cluster0
D	M	M	M	L	L	H	Cluster1
E	L	H	M	M	H	L	Cluster0

From the above observation it is clear that users C and E are in cluster 0 and users A, B and D are in cluster 1. Now the users in cluster 1 are under further investigation to reduce the false alarm rate [17, 18]. The false alarm rate refers to the check which is performed to verify whether any illegal user fall under genuine user case.

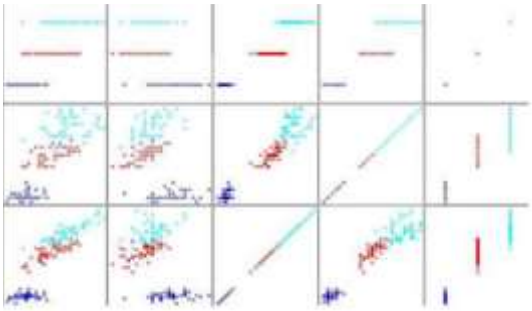

Fig 3.1: Cluster Distribution Plot

Step 5: Training Model Using Classification

Now when the complete data is clustered into two groups and in order to reduce the false alarm rate and for the detection of web crime we will pay attention to the datasets belonging to the cluster 1. We train the model using classification technique.

4. RESULTS

In every model, the accuracy plays an important role in the acceptance of that model for the application. Table III shows main results of the implementation of the user data described in table 3.1. the accuracy of the clustering comes out to be 94.75% and 5.28% comes under false alarm rate and the class is set to each user about what type of crime it is.

Table 4.1: Final Output

Users	A1	A2	A3	A4	A5	A6	Cluster	Class	Remarks
A	H	L	H	M	H	M	Cluster1	Soft	Alert
B	M	H	L	M	L	H	Cluster1	None	Genuine
C	H	H	M	H	L	M	Cluster0	None	Genuine
D	M	M	M	L	L	H	Cluster1	Hard	Crime
E	L	H	M	M	H	L	Cluster0	None	Genuine

5. CONCLUSION

The main focus of this paper is to Identify Crime Detection; this paper describes an important domain that has many problems relevant to other data mining research. It has documented the development and evaluation in the data mining for analyzing what type of crime and class it belongs. It also researches the scalable algorithm for constructing patterns of data using clustering algorithm. K-means clustering algorithm is used and then data is classified to obtain crime, none and genuine users. It efficiently detects the false rate anomalies and the types of classes in crime.

REFERENCES

[1] Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. IEEE access, 9, 70080-70094.

[2] Shah, N., Bhagat, N., & Shah, M. (2021). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Visual Computing for Industry, Biomedicine, and Art, 4(1), 9.

[3] Jenga, K., Catal, C., & Kar, G. (2023). Machine learning in crime prediction. Journal of Ambient Intelligence and Humanized Computing, 14(3), 2887-2913.

[4] Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime prediction using machine learning and deep learning: A systematic review and future directions. IEEE Access, 11, 60153-60170.

[5] Tamir, A., Watson, E., Willett, B., Hasan, Q., & Yuan, J. S. (2021). Crime prediction and forecasting using machine learning algorithms. International Journal of Computer Science and Information Technologies, 12(2), 26-33.

[6] Wu, S., Wang, C., Cao, H., & Jia, X. (2020). Crime prediction using data mining and machine learning. In The 8th International Conference on Computer Engineering and Networks (CENet2018) (pp. 360-375). Springer International Publishing.

[7] Wawrzyniak, Z. M., Jankowski, S., Szczechla, E., Szymański, Z., Pytlak, R., Michalak, P., & Borowik, G. (2018, December). Data-driven models in machine learning for crime prediction. In 2018 26th International Conference on Systems Engineering (ICSEng) (pp. 1-8). IEEE.

[8] Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021). An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: an ensemble approach. Ieee Access, 9, 67488-67500.

[9] Llaha, O. (2020, September). Crime analysis and prediction using machine learning. In 2020 43rd international convention on information, communication and electronic technology (MIPRO) (pp. 496-501). IEEE.

[10] Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022). Interpretable machine learning models for crime prediction. Computers, Environment and Urban Systems, 94, 101789.

[11] Mahmud, S., Nuha, M., & Sattar, A. (2021). Crime rate prediction using machine learning and data mining. In Soft Computing Techniques and Applications: Proceeding of the International Conference on Computing and Communication (IC3 2020) (pp. 59-69). Springer Singapore.

[12] Gahalot, A., Dhiman, S., & Chouhan, L. (2020, February). Crime prediction and analysis. In 2nd International Conference on Data, Engineering and Applications (IDEA) (pp. 1-6). IEEE.

[13] Reier Forradellas, R. F., Náñez Alonso, S. L., Jorge-Vazquez, J., & Rodriguez, M. L. (2020). Applied machine learning in social sciences: neural networks and crime prediction. Social Sciences, 10(1), 4.

[14] Chun, S. A., Avinash Paturu, V., Yuan, S., Pathak, R., Atluri, V., & R. Adam, N. (2019, June). Crime prediction model using deep neural networks. In Proceedings of the 20th Annual International Conference on digital government research (pp. 512-514).

[15] Wheeler, A. P., & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. Journal of Quantitative Criminology, 37, 445-480.

[16] Saraiva, M., Matijošaitienė, I., Mishra, S., & Amante, A. (2022). Crime prediction and monitoring in porto, portugal, using machine learning, spatial and text analytics. ISPRS International Journal of Geo-Information, 11(7), 400.

[17] Shukla, A., Katal, A., Raghuvanshi, S., & Sharma, S. (2021, June). Criminal Combat: Crime Analysis and Prediction Using Machine Learning. In 2021 International Conference on Intelligent Technologies (CONIT) (pp. 1-5). IEEE.

[18] Zhang, X., Liu, L., Xiao, L., & Ji, J. (2020). Comparison of machine learning algorithms for predicting crime hotspots. IEEE access, 8, 181302-181310.