CRIME ANALYSIS MAPPING, INTRUSION DETECTION - USING DATA MINING

Gundlapalli Harshitha[1], K. Sravan Kumar[2]

[1]M.Tech Student, Department of computer science engineering, Vinuthna Institute of Technology & Science, Hasanparthy (Mdl), Warangal, Dist, Hasanparthy, Telangana 506371

[2]Assistant professor, Department of computer science engineering, Vinuthna Institute of Technology & Science, Hasanparthy (Mdl), Warangal, Dist, Hasanparthy, Telangana 506371

Abstract

Data Mining is an important part of the process of Investigating Crimes.Earlier research publications have made reference to a wide variety of methods, a few examples of which include the virtual identifier, the elliptic curve algorithm, and the binary treealgorithm. Strategy for removing unwanted branches, support vector machines, and a priori VID is used to determine the relationship that exists between record and vid. In order to aid the fuzzy classification models, the apriori algorithm was developed. method, and around six hundred seconds are required to discover it. an assault using bombs sent in the mail. Within the scope of this study report, we found theAnalysis of crime maps based on the KNN algorithm (K – Nearest Neighbour) and ANN (Artificial Neural Network) are two examples of these kind of algorithms. help make this procedure easier to understand. Mapping of criminal activity is performed, and Obtainable thanks to the Office of Community-Oriented Policing's financial support Providers of (COPS). Research that is based on evidence is helpful in investigating the criminal acts. We determine the crime rate based on the following: the prior data by using various data mining methods. Crime Analysis employs quantitative and qualitative data in conjunction with analytical methods, with the purpose of solving the situations. The mapping of criminal activity is an important tool for ensuring public safety. crucial scientific field that must get attention. We are able to identify the areas with the highest rate of criminal activity with the assistance of data mining methods. Within the framework of Crime Analysis Mapping, we In order to bring down the number of reported crimes, please proceed as follows:

1) Collect criminal data 2) Group data 3) Clustering4) Making projections based on the facts.

I.INTRODUCTION

There is a wide variety of criminal activity that may take place, such as armed robbery, the theft of motor vehicles, and many more. When there is a higher overall crime rate, the investigative procedure becomes both more time consuming and difficult.

The resolution of even the most complex legal matters is often aided by the use of information extraction techniques. Crime analysis combined with crime mapping is one of the most effective ways. Together, crime analysis and crime mapping contribute to a better knowledge of the principles and procedures of crime analysis, which in turn aids the police and contributes to the reduction and prevention of criminal activity and criminal disturbances. The Office of Proactive Policing Service is responsible for both the conduct and funding of crime mapping (COPS). Research that is based on evidence is helpful in assessing the crimes. Using methods from data mining, we determine the crime rate by calculating it based on the data from before. Crime analysis is a method that helps solve

crimes by making use of quantitative and qualitative data as well as analytical tools.

The mapping of criminal activity is an important study topic to focus on for reasons having to do with public safety. Through the use of data mining algorithms, we are able to pinpoint the areas with the most potential for criminal activity.

II.EXISTING WORK

The number of reported crimes is rising on a daily basis, and people all over the globe are scrambling to find ways to control the crime rate and to make progress on specific instances. The vast majority of individuals are attempting to collect relevant data for use in the future. Errors resulting from human behaviour are always a possibility. There are a variety of offences that fall within the jurisdiction of law enforcement, including traffic infractions, sexual offences, theft, acts of violence, arson, offences related to gangs or drugs, and cybercrime. Amongst each of them, many criminal data mining strategies are offered, such as entity extraction, clustering algorithms, and Association rule mining. Crime hotspots may be used to determine which areas are prone to higher rates of criminal activity. There is a need for patrol in these high-risk regions.

The application for data mining contributes to a significant decrease in the number of reported crimes [1].

As a result of the dramatic rise in the use of networks, security has emerged as one of the most pressing concerns about these systems. The following are some of the reasons why data mining is utilised in malicious network intrusions:

1. It is suited to identify the disregarded and concealed knowledge at any point in time.
2. To handle enormous volumes of data.
3. It is suitable to detect the ignored as well as hidden information.

The Intrusion Detection System (IDS) is what's used to find problems that are connected to intrusions on networks. Machine learning is the process of designing and developing algorithms in a manner that enables computers to acquire information about the information that is provided to the machine. This process is referred to as "machine learning." In fields such as bioinformatics, machine learning is used to detect the pattern match in DNA and to check for data connected to genes.

The primary responsibility of the Intrusion Detection System is to identify both genuine threats and false positives [2].

It assists in identifying legitimate and fraudulently authentic users, which is beneficial for the maintenance of users' privacy. In recent times, there has been an increase in both false alarms and intrusions, and the methods they are using are significantly different from the conventional ones. The problem of vulnerability scanning may be handled by using data mining. For instance, the United States Army use it for the management of restrictions in military systems while operating in tactical contexts [3].

One of the most popular types of assaults carried out against websites is the distributed denial of service attack. DoS assaults may be thwarted utilising the information that is gleaned via intrusion detection, which aids in determining what kind of network activity is taking place.

Both misuse detection, which relies on an exact template matching, and anomaly - based, which needs further training in relation to artificial intelligence, are offered here as techniques for detecting intrusions. misuse detection is based on an exact pattern match.

A Fuzzy Intrusion Recognition Engine (FIRE) is an Anomaly Intrusion Detection System (IDS) that uses fuzzy methods to identify hostile websites as being untrustworthy. In this case, a three-dimensional packet count with a 15-minute interval is utilised to locate the normal network connections and to make an attempt to identify any intrusions that may have occurred at that particular moment in time [4].

A computer's health is comparable to that of a human's in that it requires protection from harmful elements. The usage of fuzzy cognitive maps and fuzzy rules both contribute to and are utilised for the acquisition of causal knowledge.

As the number of malicious acts committed using computers continues to rise, so does the urgency with which we must safeguard our information. As part of the whole process of intrusion detection, the intelligent intrusion detection system is also built.

The values on a fuzzy cognitive map shift around from time to time, and there are ties of causation seen between nodes that are used to represent the directed edges [5].

In this, they identified patterns of criminal behaviour by using clustering algorithms. Clusters is a term that refers to the requirement for determining the location and nature of a crime at a certain moment in time within a geographic region. To locate the location of the plot, we may utilise a map. The most difficult obstacle to overcome is free text areas.

Converting the free text fields into data might be a challenging task; nevertheless, the Kmeans approach is used in this study in order to accomplish this goal.

This method allows for the extraction of operational data as well as their subsequent transformation into another form. When this is done, it is considerably simpler for the investigators to uncover the patterns of criminal behaviour that they need to identify the frauds [6].

It is critical to have an intrusion detection system in place so that you can differentiate between aberrant and typical patterns of behaviour. In order to identify intrusions, we make use of fuzzy logic. This system combined with fuzzy logic employs association rule mining, the quickest way \sis to use prefix trees. In addition to this, it assists in the process of processing and extracting the data by making use of the date and source host information. Utilizing the traffic in the audit logs, we are able to locate the assaults. TCP header is used for the purpose of more effectively improving efficiency. The pruning stage is used to cut down on the amount of time needed to operate the system while simultaneously improving its level of precision [7].

The majority of individuals have a skewed understanding of how data mining works, as well as its applications and when and when it might be employed. The protection of one's privacy is only one of the numerous advantages offered by data

mining. Every every transaction that takes place in the modern world is documented and filed away in a certain location.

It is imperative that confidentiality be preserved for these kinds of information.

Data accumulated over years may be found in warehouses. The fundamental challenge is the most important issue at hand; all of the data must be saved in a single location, and the time and resources needed to retrieve such a massive quantity of data would be quite expensive. Therefore, this has developed into the primary concern.

Datamining was able to overcome this problem by splitting the data in both a horizontal and vertical fashion. When the data is stored using the horizontal portioning method, it is a great deal simpler to obtain the data [8]. In today's world, the primary responsibility of any company or organisation is to protect its computer networks by using intrusion detection systems. This document provides readers with a variety of algorithms from which to choose and implement.

The fields of machine learning and pattern recognition both benefit from the use of decision trees. ID3 is an algorithm that is based on machine learning and it identifies the branches of a tree by using the root of the tree as the identifying feature. Cross-validation tests have been carried out in order to determine the patterns, and comparisons have been made between the various kinds of algorithms in order to establish the characteristics and preserve the efficiency [9].

Utilizing the system logs, we are able to conduct an investigation into the breach.

Misuse of either private or public information may result in an intrusion. Through the use of Intrusion Systems, we are able to recognise illegal users. Because the logs could take up a lot of space in the system, we need to come up with ways to manage them that are both more flexible and less expensive. In order to evaluate the speed and scalability, support vector machine intrusion detection is put through its paces. A typical assault is constructed using data from 22 distinct examples in order to determine a pattern in this data. The training of the neural network systems is what has to be determined as the aim. The fact that neural networks have been employed in a variety of IDSs may be shown by the fact that this technology is used for several categorization categories [10].

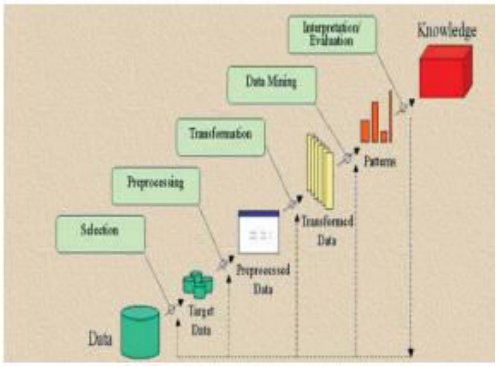
III. PROPOSED APPROACH

Applying methodologies, crime mapping aids in the reduction and prevention of crimes and criminal disorders, as well as in the comprehension of the ideas and practise of crime analysis, which is of assistance to the police. We might make use of data mining technologies that incorporate artificial neural networks and knowledge discovery in databases (KDD and ANN, respectively).

We gather the data first from law enforcement and make an effort to acquire as much information as possible, including the person's name, height, age, sex, fingerprint details, and pattern identification number for instances that are similar to one another. After we have

obtained the information, we will immediately begin processing the data.

Along with the data that is essential, we receive a lot of data that is not necessary. However, before we begin processing the data using the strategies and tools for data gathering, we need to determine which data are superfluous and then get rid of those types of data in order to lessen or completely prevent the confusion. In order to determine the pattern in the crime data, we make use of the SAM tool. We possess supervised information and unsupervised information in this case. These are the two categories of data. We use the information that includes every specific detail regarding the case, and we make an effort to Use of licenced materials that is authorised is restricted to: Southern Queensland University, sometimes known simply as USQ. Retrieved in the month of August

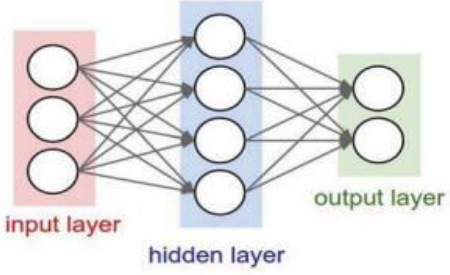


Apart from the data that is essential, we receive a lot of data that is not necessary. However, before we begin processing the data using the methods and techniques for data gathering, we need to determine which data are superfluous and then get rid of those types of data in order to lessen or completely prevent the confusion. In order to determine the pattern in the crime data, we make use of the SAM tool. We have supervised data and unsupervised data in this case.

These are the two categories of data. We start with the data that has all of the specifics of the case, and utilising this supervised data as a basis for our training, we attempt to solve the other instances. We focus primarily on collecting information on traits such as eye colour, fingerprints characteristics, features, size, and any other factors that may be relevant.

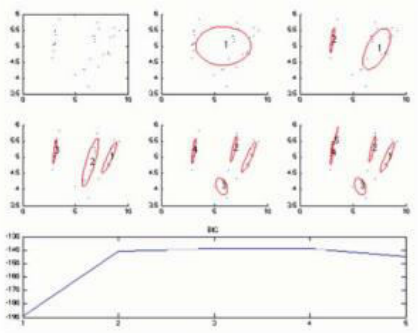
In its most fundamental form, neural networks are composed of three components: the engineering, often known as the model; the learning calculation; and the enactment capabilities. "...store, perceive, and collaborative manner recover arguments in favour or database sections; to require control of complex mathematical augmentation issues; to channel clamour from assessment knowledge; to control not well characterised issues; in quick summary, to reassess tested capabilities when we do not have a clue about just the type of something like the capacities."

Neural systems are customised or "prepared" to "...store, comprehend, and collaborative manner recover examples or metadata segments; to take care of complex mathematical enhancement issues; to channel clamour from prediction knowledge; to control not



For this, we utilize the KDD which is a learning revelation from information. This procedure includes the extraction of the fascinating data which implies the information ought to be nonrehashing, understood (past obscure and right now it is helpful) from a colossal measure of the information. Likewise, this includes the example coordinating, collecting the data and business knowledge and so forth. With this device, adequate information mining will be performed.

The center of the KDD is the information mining. The means associated with the KDD are as per the following: data cleaning, data putting away, crime related information and again design, and evaluation. At long last with these means we will get the profitable data.



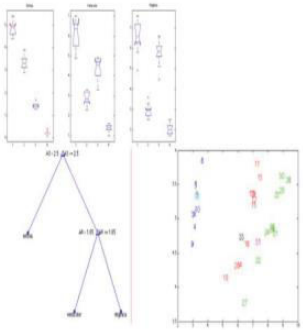
In addition to the tool, we may have to perform a thorough inspection on the information to assure that the data that was obtained is accurate. In light of the surrounding factors, informational reliability issues manifest themselves in some of the instances that are investigated.

They are redundant information, noise, insufficient information, and inconsistent data. The knowledge is inconsistent. These factors result in the production of bad information, which need to be discarded before the documentation is dissected.

IV. EXPERIMENTS

The instruction is supplemented by an indication of the class of the inspections and is subject to supervision.

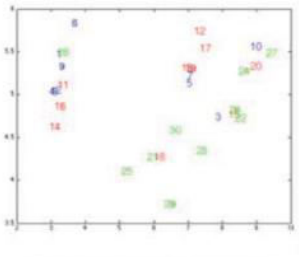
The training set is used to guide the classification of newly acquired data. Making use of the training phase, we locate patterns that are a match.



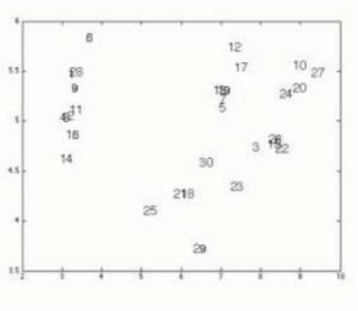
Learning without a teacher present (clustering)

It is not known what the class labels of the training data are.

Given a collection of measures, inspections, and other such things, with the intention of determining whether or not there are classes or groupings in the data,



The order is no longer meaningful



There are still patterns

They would collect the data from the various police offices with the aid of these equipment, and then we will choose the various information for the purposes of evaluating and compiling the material. After going through these steps of information examination, this data will then be updated at the communication centre point, where the various police factions will have access to it. This data will actually prove any oddities in the knowledge and provide a clear connection between the data, with the goal of preventing and protecting the general public from illegal activity.

V. CONCLUSIONS

The information concerning the illegal activity will be sent to the device that digs up some information for further investigation with the help of these devices, and the results of the investigation will afterwards be recorded for two different models. We will keep a deliberate distance from the distinction in the accomplishment with the assistance of the SAM instrument/tools, and after that, the resulting knowledge will be utilised for the discovering of the connections between those. In this manner, we will lower the number of false positives and false negatives in the area of the interruption recognition framework by applying data extraction in the field of criminal activity data examination.

REFERENCES

[1] Chen, Hsinchun, et al. "Crime data mining: a general framework and some examples." computer 37.4 (2004): 50-56.

[2] Ektefa, Mohammadreza, et al. "Intrusion detection using data mining techniques." Information Retrieval & Knowledge Management, (CAMP), 2010 International Conference on. IEEE, 2010.

[3] Clifton, Chris, and Gary Gengo. "Developing custom intrusion detection filters using data mining." MILCOM 2000. 21st Century Military Communications Conference Proceedings. Vol. 1. IEEE, 2000.

[4] Dickerson, John E., and Julie A. Dickerson. "Fuzzy network profiling for intrusion detection." Fuzzy Information Processing Society, 2000. NAFIPS. 19th International Conference of the North American. IEEE, 2000.

[5] Siraj, Ambareen, Susan M. Bridges, and Rayford B. Vaughn. "Fuzzy cognitive maps for decision support in an intelligent intrusion detection system." IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th. Vol. 4. IEEE, 2001.

[6] Nath, Shyam Varan. "Crime pattern detection using data mining." Web intelligence and intelligent agent technologyworkshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on. IEEE, 2006.

[7] Florez, German, S. A. Bridges, and Rayford B. Vaughn. "An improved algorithm for fuzzy data mining for intrusion detection." Fuzzy Information Processing Society, 2002. Proceedings. NAFIPS. 2002 Annual Meeting of the North American. IEEE, 2002.

[8] Panda, Mrutyunjaya, and Manas Ranjan Patra. "A comparative study of data mining algorithms for network intrusion detection." Emerging Trends in Engineering and Technology, 2008. ICETET'08. First International Conference on. IEEE, 2008.

[9] Vaidya, Jaideep, and Chris Clifton. "Privacy-preserving data mining: Why, how, and when." IEEE Security & Privacy 2.6 (2004): 19-27.

[10] Mukkamala, Srinivas, Guadalupe Janoski, and Andrew Sung. "Intrusion detection using neural networks and support vector machines." Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on. Vol. 2. IEEE, 200

Author Profile

K Shravan kumar, graduated from Vaagdevi College of Engineering, post graduated from aurora engineering college and working as assistant professor in Vinuthna Institute of Technology & Science since 2016 to present.