Data Mining Based Crime Analysis Mapping and Intrusion Detection

Biswajit Panja
Eastern Michigan University

Priyanka Meharia
Eastern Michigan University

Keerthi Mannem
Eastern Michigan University

Data Mining plays a key role in Crime Analysis. There are many different algorithms mentioned in previous research papers, among them are the virtual identifier, pruning strategy, support vector machines, and apriori algorithms. VID is to find relation between record and vid. The apriori algorithm helps the fuzzy association rules algorithm and it takes around six hundred seconds to detect a mail bomb attack. In this research paper, we identified Crime mapping analysis based on KNN (K – Nearest Neighbor) and ANN (Artificial Neural Network) algorithms to simplify this process. Crime Mapping is conducted and Funded by the Office of Community Oriented Policing Services (COPS). Evidence based research helps in analyzing the crimes. We calculate the crime rate based on the previous data using data mining techniques. Crime Analysis uses quantitative and qualitative data in combination with analytic techniques in resolving the cases. For public safety purposes, the crime mapping is an essential research area to concentrate on. We can identity the most frequently crime occurring zones with the help of data mining techniques. In Crime Analysis Mapping, we follow the following steps in order to reduce the crime rate:

1) *Collect crime data*
2) *Group data*
3) *Clustering*
4) *Forecasting the data.*

Crime Analysis with crime mapping helps in understanding the concepts and practice of Crime Analysis in assisting police and helps in reduction and prevention of crimes and crime disorders.

Keywords: data mining, data security, user privacy, supervised learning, unsupervised learning

INTRODUCTION

Crimes are one of the most predominant problems that is happening in most of the urban areas in the world. There are a lot of different types of crimes that happen, including robbery, theft of vehicles, etc. As crime increases, the investigation process gets longer and more complicated.

The use of information mining methods helps in resolving most complicated criminal cases. One of the best methods is crime analysis with crime mapping. Crime analysis with crime mapping helps in understanding the concepts and practices of crime analysis in assisting police and helps in the reduction and prevention of crimes and crime disorders.

Crime mapping is conducted and funded by the Office of Community Oriented Policing Services (COPS). Evidence based research helps in analyzing the crimes. We calculate the crime rate based on the previous data using data mining techniques. Crime analysis uses quantitative and qualitative data and analytic techniques in resolving the cases.

For public safety purposes, the crime mapping is an essential research area to concentrate on. We can identify the highest risk crime zones with the help of data mining techniques.

EXISTING WORK

Crime has been increasing day by day and everyone in the world is trying to figure out how to manage the crime rate and to work on certain cases, most of the people are trying to store the data for future reference. Human errors can occur at any point of time. There are different types of crimes law enforcement levels, such as traffic violations, sex crime, theft, violent crime, arson, gang/drug offenses, cybercrime. Different crime data mining techniques are proposed among each of them including entity extraction, clustering techniques, Association rule mining. Crime zones can be identified by occurrence of crime, by using hotspots. Patrol is needed at these hotspot areas. The data mining tool helps in reducing the crime rate drastically.

Security is considered to be a major issue in networks as the usage of networks has increased drastically. Data mining is used in network intrusions because of the following reasons:

- To process huge amounts of data.
- It is suitable to detect the ignored and hidden information at any point of time.

The Intrusion Detection System is applied to detect intrusion network related issues. Machine Learning is to deal with design and development of algorithms and in a way to allow computers to learn about the data that is fed to the machine. Machine learning is applied in areas like bioinformatics, to find the pattern match in DNA and to check for gene related data. Detection of attacks and false alarms is the main task of the Intrusion Detection System.

It helps to identify authentic and falsely authentic users for maintaining privacy. False alarms and intrusion is more in recent days, and the techniques they are following is more different than the usual ones. Using data mining, intrusion detection can be resolved. For example, it is used by the US Army for tactical environments to handle constraints in military systems.

Distributed denial of service attacks is one of the most common attacks on internet sites. Intrusion detection helps in identifying the network related activity and using this we can provide security against DoS attacks. There are two types of intrusion detection methods available here: misuse detection which is based on exact pattern match, and anomaly detection which requires more training related to artificial intelligence. Fuzzy intrusion recognition engine is an anomaly IDS which identifies malicious sites which are not trustworthy using fuzzy systems. Here, 3-D packet count with a 15-minute interval is used to find the regular network connections and try to indicate the intrusions, if any, at that point of time.

The health of a computer is analogous to that of human health: it needs protection. Fuzzy cognitive maps and fuzzy rules are used for and support causal knowledge acquisition. Computer crimes are increasing day by day, and with it the need to protect our data has also increased. The intelligence intrusion detection system is also developed as part of intrusion detection. Fuzzy cognitive map values are changed from time to time and there are causality links between nodes that represent the directed edges.

In this, they used clustering techniques to identify crime patterns. In a geographical area, the need to identify the crime at a point in time is known as clustering. We can use a map to identify the plot. The largest challenge is with free text fields. It is difficult to convert the free text fields into data, but the K-means technique is used for this purpose in this paper. Operational data can be extracted and transformed to another form using this technique. By doing this, it is much easier to find out the crime patterns for the detectives to identify the frauds.

To identify the difference between the abnormal and normal activities, the intrusion detection system is very important. We use fuzzy logic in intrusion detection. This system along with fuzzy logic uses association rule mining, the fastest technique is to use prefix trees. It also helps in transforming and extracting the data by using timestamp and source host details. We can identify the attacks by using traffic in audit logs. TCP header is used for more improved efficiency. The pruning step is used to reduce the running time and to increase the accuracy of the system.

Most people misunderstand how to use data mining and where and when it can be used. Data mining provides many benefits, one of which being privacy. In the present world, each transaction is recorded and stored in some particular place. Privacy needs to be maintained for these types of data. Warehouses are used to store years of data. The main problem here is the central problem, where all the data needs to be stored in a single place and required to retrieve this large bulk amount of data would cost a lot. So, this has become the main issue. Datamining resolved this issue by using horizontal and vertical partitioning of data. It is far easier to retrieve the data when stored in the horizontal portioning way.

In modern days, securing the computer systems using intrusion detection systems has been the main task for any organization. There are different types of algorithms to work with in this paper. Decision trees help in areas like machine learning and pattern recognition. ID3 is a machine learning based algorithm, which uses the root of the tree as the attribute to identify the branches. Cross validation tests have been performed to identify the patterns and comparisons between the different types of algorithms to identify the characteristics and maintain the effectiveness.

We can analyze the intrusion using logs from the system. Intrusion happens due to misuse of private or public data. We can identify unauthorized users using Intrusion Systems. We need more flexible, cost-effective systems for handling the logs since they may occupy a lot of space in the system. Support vector machine intrusion detection are used to test the speed and scalability. A normal attack is based on 22 different cases and formed to identify a pattern in this. The goal is to identify the training of the neural network systems. This system is used for multi classification categories which proves that neural networks has been used in many IDs.

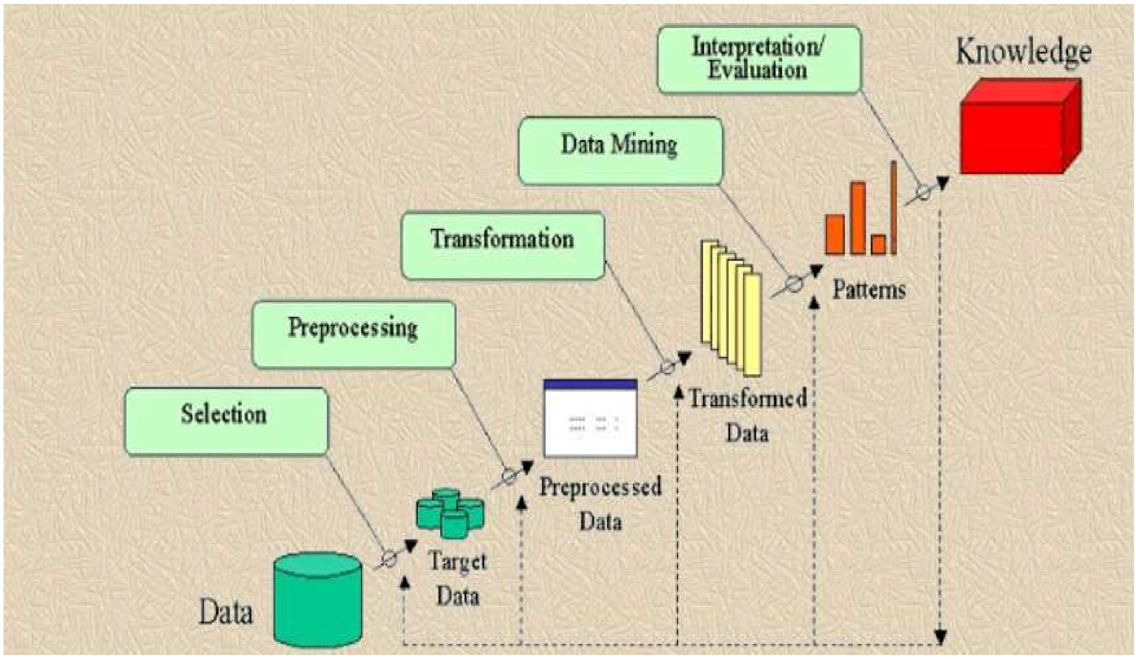
PROPOSED APPROACH

Crime Mapping helps in understanding the concepts and practice of Crime Analysis in assisting police and helps in reduction and prevention of crimes and crime disorders using data mining tools. We can use data mining tools involved using ANN (Artificial Neural Networks) and KDD (Knowledge Discovery in Databases).

We collect the data from police department and try to get each and every detail, like the person’s name, height, age, sex, fingerprint details, and pattern identification number for similar types of cases. Once we get the information, we start to process the data.

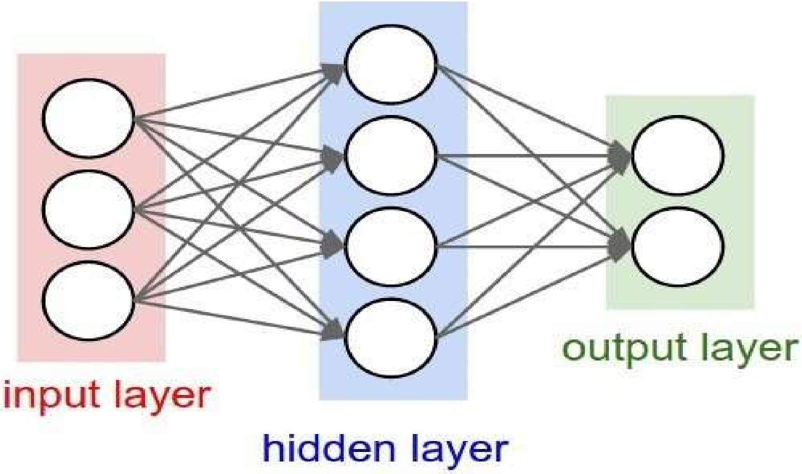
We get a lot of unnecessary data along with the required data. But before we start processing the data using data mining techniques and tools, we need to identify unnecessary data and remove those kinds of data to reduce or to avoid the confusion. We use the SAM tool to identify the pattern in the crime data. Here we have two classifications of data: supervised and unsupervised data. We take the data that has all the details about the case and we try to solve the other cases by training using this supervised data. We mainly collect the attributes information, like eye color, fingerprint details, characteristics, dimensions, or other features.

FIGURE 1
DATA PROCESSING STEPS



Neural systems basically include three pieces: the engineering, or model; the learning calculation, and the enactment capacities. Neural systems are customized or "prepared" to "...store, perceive, and cooperatively recover examples or database sections; to take care of combinatorial enhancement issues; to channel clamor from estimation information; to control not well characterized issues; in rundown, to evaluate tested capacities when we don't have a clue about the type of the capacities."

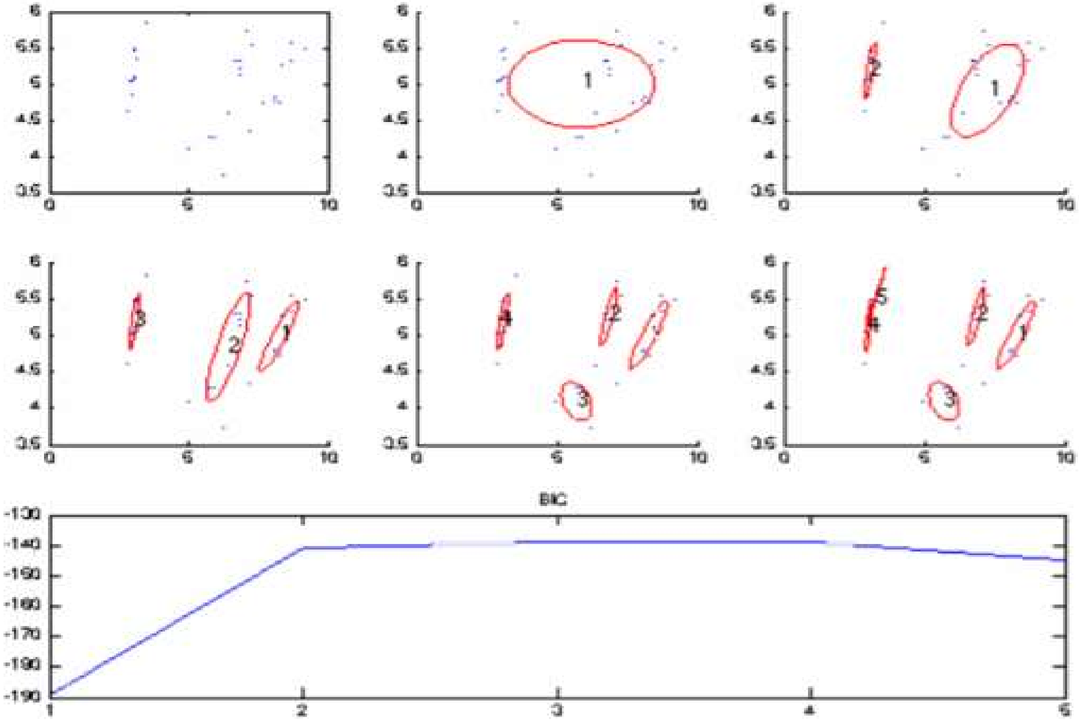
FIGURE 2
NEURAL NETWORK INPUT AND OUTPUT



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 non-rehashing, 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.

FIGURE 3
DATA CLASSIFICATION

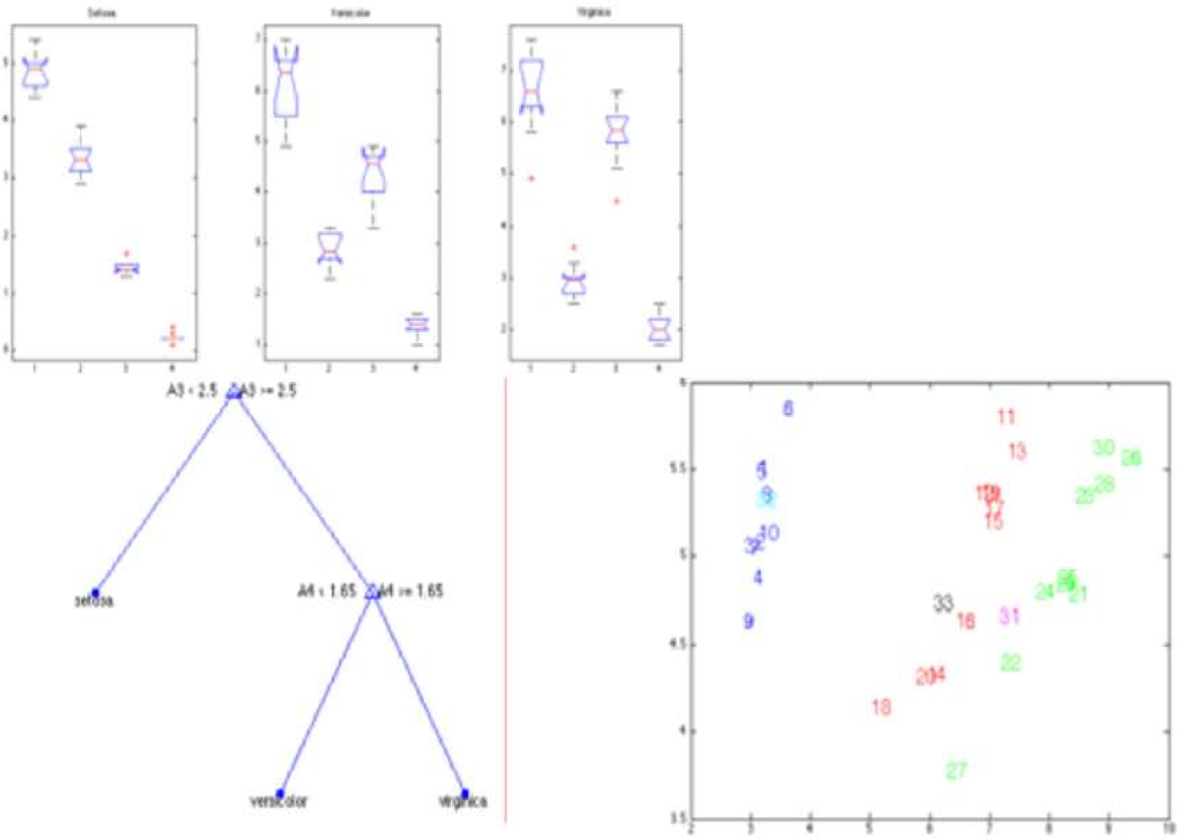


Alongside the instrument we may need to check the information quality to ensure that discovered information is decent. A portion of the cases of the information quality issues emerges in view of the accompanying things. They are duplicate information, noise, incomplete information, and inconsistency of the information. These things prompt terrible information, which should be expelled before we dissect the information.

EXPERIMENTS

Supervision: the training is accompanied by indicating the class of the observations. New data is classified based on the training set. We identify matched pattern by using the training set.

FIGURE 4
DATA CLASSIFICATION AND PATTERN RECOGNITION



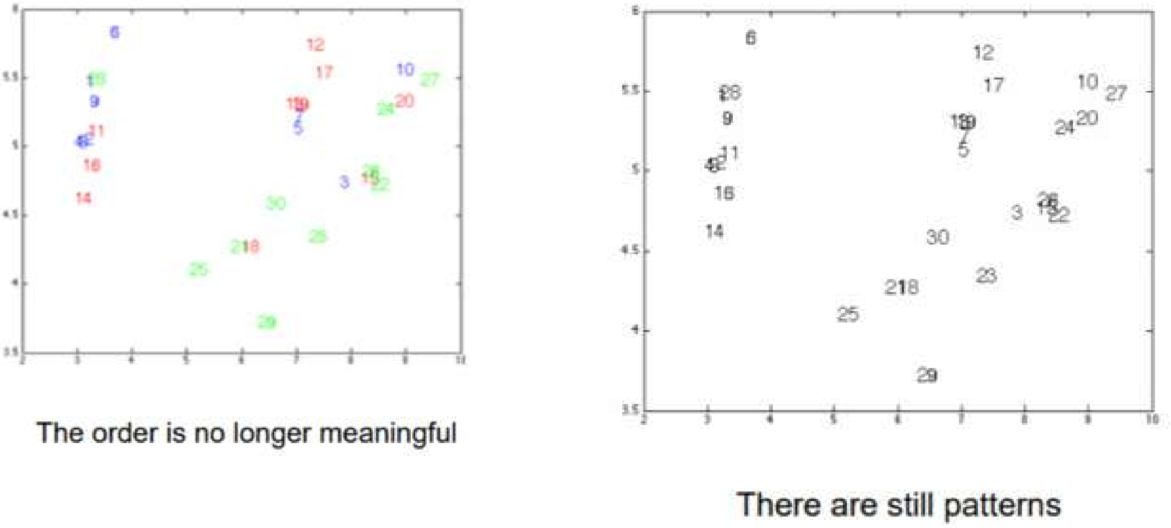
5.1,3.5,1.4,0.1
4.9,3.0,1.4,0.2
4.7,3.2,1.3,0.2
4.6,3.1,1.5,0.2
5.0,3.6,1.4,0.2
5.4,3.9,1.7,0.4
4.6,3.4,1.4,0.3
5.0,3.4,1.5,0.2
4.4,2.8,1.4,0.2
4.9,3.1,1.5,0.1
7.0,3.2,4.7,1.4
6.4,3.2,4.5,1.5
6.9,3.1,4.9,1.5
5.5,2.3,4.1,1.3
6.5,2.8,4.6,1.5
5.7,2.8,4.5,1.3
6.3,3.3,4.7,1.6

4.9,2.4,3.3,1.0
6.6,2.9,4.6,1.3
5.2,2.7,3.9,1.4
6.3,3.3,6.0,2.5
5.8,2.7,5.1,1.9
7.1,3.1,5.9,2.1
6.3,2.9,5.6,1.8
6.5,3.0,5.8,2.2
7.6,3.0,6.6,2.1
4.9,2.5,4.5,1.7
7.3,2.9,4.3,1.8
6.7,2.5,5.8,1.8
7.2,3.6,6.1,2.5

UNSUPERVISED LEARNING

The class labels of training data is unknown. Given a set of measurements, observations, etc. with the aim of establishing the existence of classes or clusters in the data.

FIGURE 5
LEARNING AND DATA CLASSIFICATION



The order is no longer meaningful

There are still patterns

5.0,3.6,1.4,0.2
4.9,3.0,1.4,0.2
6.3,2.9,5.6,1.8
4.7,3.2,1.3,0.2
6.5,2.8,4.6,1.5
5.4,3.9,1.7,0.4
6.3,3.3,4.7,1.6
4.6,3.5,1.4,0.3
5.0,3.4,1.5,0.2
7.2,3.6,6.3,2.5
4.9,3.1,1.5,0.1
7.0,3.2,4.8,1.4
6.4,3.2,4.5,1.5
4.4,2.1,1.4,0.2
6.7,2.4,5.8,1.8

4.6,3.1,1.5,0.2
6.9,3.1,4.9,1.5
5.5,2.3,4.0,1.3
6.6,2.9,4.6,1.3
7.3,2.9,6.3,1.8
5.2,2.7,3.9,1.4
6.3,3.3,6.0,2.5
5.8,2.7,5.1,1.9
7.1,3.0,5.9,2.1
4.9,2.4,3.3,1.0
6.5,3.0,5.8,2.2
7.6,3.0,6.6,2.1
5.1,3.5,1.2,0.2
4.9,2.5,4.1,1.7
5.7,2.8,4.5,1.3

With the assistance of these devices we will gather the data from the police divisions, and we will choose the different information for testing and preparing information. These information subsequent to experiencing these information mining steps, it will be refreshed to the information center point where the police divisions will have an entrance to that information which will substantiate the awkwardness in the information and give clear connection between the information to keep and shield the general population from wrongdoing.

CONCLUSIONS

With the assistance of these devices, the wrongdoing information will be nourished to the information digging device for investigation and afterward comes about for two unique models will be recorded. With the assistance of the SAM instrument/tools, we will maintain a strategic distance from the distinction in the outcome and after that the subsequent information will be utilized for the finding the relations amongst those et cetera. Along these lines we will lessen false positives and false negatives in the field of the interruption identification framework utilizing the information mining in the field of wrongdoing information examination.

REFERENCES

Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: A general framework and some examples. *Computer*, *37*(4), 50–56. https://doi.org/10.1109/mc.2004.1297301

Ektefa, M., Memar, S., Sidi, F., & Affendey, L.S. (2010). Intrusion detection using data mining techniques. *Information Retrieval & Knowledge Management, (CAMP) 2010 International Conference*. IEEE.

Clifton, C., & Gengo, G. (2000). Developing custom intrusion detection filters using data mining. MILCOM 2000. *21st Century Military Communications Conference Proceedings* (Vol. 1). IEEE.

Dickerson, J.E., & Dickerson, J.A. (2000). Fuzzy network profiling for intrusion detection. Fuzzy Information Processing Society, NAFIPS. *19th International Conference of the North American*. IEEE.

Siraj, A., Bridges, S.M., & Vaughn, R.B. (2001). Fuzzy cognitive maps for decision support in an intelligent intrusion detection system. *IFSA World Congress and 20th NAFIPS International Conference*. Joint 9th. Vol. 4. IEEE.

Nath, S.V. (2006). Crime pattern detection using data mining. Web intelligence and intelligent agent technology workshops, wi-iat 2006 workshops. *2006 ieee/wic/acm International Conference*. IEEE.

Florez, G., Bridges, S.A., & Vaughn, R.B. (2002). An improved algorithm for fuzzy data mining for intrusion detection. Fuzzy Information Processing Society. *Proceedings. NAFIPS*. 2002 Annual Meeting of the North American. IEEE

Panda, M., & Patra, M.R. (2008). A comparative study of data mining algorithms for network intrusion detection. Emerging Trends in Engineering and Technology. ICETET'08. *First International Conference on IEEE*.

Vaidya, J., & Clifton, C. (2004). Privacy-preserving data mining: why, how, and when. *IEEE Security and Privacy Magazine*, *2*(6), 19–27. https://doi.org/10.1109/msp.2004.108

Mukkamala, S., Janoski, G., & Sung, A. (2002). 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.