Technoarete Transactions on Intelligent Data Mining and Knowledge Discovery
Vol-2, Issue-4, November 2022
e-ISSN: 2583-195X

Cybersecurity using Data Mining Techniques

Dr. Sanjiv Kumar Jain [1], Joey G. Fernando [2]

[1] Medi-Caps University, Indore, India.
[2] Central Luzon State University, Philippines.
*Corresponding Author Email: [1] sanjivkj@gmail.com

Abstract

The study is based on the cyber security using data mining in a certain manner and all the insights have been taken from reliable resources which are related to the subject matter. At the beginning of the following article, the, the introduction has been served over the following topic of the cyber security and data mining in business. Also, in the following context, several kinds of materials and methods have been selected and evaluated to bring betterment in the study. The cross-sectional research design, inductive approach, secondary data have been selected by following the qualitative method. At the beginning, the data mining and cyber security have been represented with sheer elaboration and then the importance of data mining techniques within the business has been illustrated within a certain manner. After this, different kinds of usages of data mining techniques like detecting malwares, intrusions and virus attacks have been depicted in a certain norm of data mining techniques. Later on, the following study has been evaluated with the help of proper and authentic interpretation of the rate of cyber security within a business.

Keywords

Cyber security, data mining techniques, data mining, malware.

INTRODUCTION

In the terms of cyber security within the era of technology implementation, cyber security has become an important aspect which has to be maintained in a proper manner in the cyber world of todays' time. The cyber security intends to pay attention to securing computational systems from corrupted and unauthorized entry or being otherwise damaged or made inaccessible in a certain manner. Also, information security is an extensive category that secures all informational factories, in hard copy or in a digitalised form as well. In short, cyber security is a clear application of technology, procedures and regulations to secure the systems, networks and the programs within a certain manner from the data cyber-attack [1]. Cyber security is an important aspect because it makes all types of data safeguards against all kinds of thefts, threats and loss and protects the sensitive data which are confidential from all aspects of a cyber-cell. Basically, cyber security regulates and secures all sorts of data which are related within a particular company and explorations of systems, network and technologies in a certain manner.

By using data mining techniques and strategies, the cyber security can be increased and maintained in a successive manner within a particular business. Implementing the data mining techniques, a business can be able to examine the outcomes of particular audits and fetch and unrecognized paradigms which affect a business from all aspects. Therefore a business by using data mining techniques can be able to detect the instructions, network and systems scanning and denial of services in a successive manner and also, penetrate the threats which can break the pace of cyber security [2]. Also, in order to fetch the host based attacks, the cyber security software requires examining the features which have been extracted from the programs in an immense successive manner. Henceforth it can be stated that the use of data mining in the process of cyber security might lead to an immense amount of intrusion in a certain manner.

Data mining has been typically constructed within cyber security applications which can be immensely impactful to a business in a certain manner. Also, as an application of developing the speed and quality of malfunction detention, incremental kinds of strategies can be implemented to fetch effective malware involving anomaly detection and misleading detection within a business [3]. Furthermore, data mining and cyber security can also be impactfully implemented to fetch in authorized intuitions and examine the audits outcomes to spot the abnormal patterns of business in a certain manner. Also, two types of data mining can be able to detect the intrusions which are closely related to the malware of business and the types of data mining are as follows- predictive data mining and descriptive data mining.

MATERIALS AND METHODS

In the following study, there are several types of materials and methods that have been selected and executed in a certain manner of the research work and the study has followed the stereotypical norm of a research. The core subject matter of this study is based on the concept of cyber security and its regulation by data mining within a business and the subject matter has been served by collecting proper data over the following topic of cyber security within a business in a certain manner. In the following study, all the possible aspects have been made which are related to methodology of a research work.

For the following study, there are few crucial types of materials and methods that have been selected and executed for the following research work. In this study, the cross sectional research design has been selected to evaluate the complex subject matter in a sorted manner. The reason for choosing cross sectional data of the research work is to evaluate the subject matter from various kinds of aspects. Data which has been collected for the study are secondary by nature and the data has been collected by following the qualitative method. Also, inductive approach is chosen to gain insights based on study.

RESULTS

In recent days, prioritizing data mining is escalating at a high rate as raw information can be converted into useful and essential data with the aid of data mining. Data mining helps several businesses to gain an explicit understanding of customers' needs [4]. Currently, companies are capable of accelerating existing profit, boosting relationships with customer segments, and reducing certain expenses related to business operations through appropriately utilising data mining. Every business has to maintain the security of confidential information; however, in many cases, some firms lose important data and information due to unexpected fraudulent activities. In this field, data mining can aid businesses to enhance the security level to protect necessary business information and data. In the present case an instance can be provided; often unwonted behaviour and patterns can be identified by anomaly detection which is an indispensable part of data mining. Viruses can be detected by link analysis that protects information from frauds [5]. Data mining has great importance in making malware detection appropriate.

Several needs can be noticed in cyber security and one of the most usual requirements is malware detection. Technique related to the detection of malware presence in a system is considered as malware detection [6]. Businesses can secure information in systems from the touch of hackers with the aid of malware detection. Data mining plays an essential part in intensifying overall quality of detection of malware. Speed related to malware detection is also ameliorated by data mining procedure. Any type of intrusions can easily be detected by data mining and it consists of necessity in interpreting outcomes of audit. Analysis of audit results subsequently aids to recognise anomalous patterns. There are various kinds of malicious intrusions such as intrusions into operating systems, web clients, servers, databases, and networks.

By examining security logs and databases with "data mining strategies" can be able to help a business to detect the malware, systems and network mansions, "insider attacks" and several "security threats". There are some strategies which can even accurately assume attacks and several other security threats. There are few strategies that can even apparently assume the attacks and detect zero threats within a few times. Also, there are some key strategies and uses of data mining in network and endpoint encryption through which all types of data can be secured for a business a certain way. Data mining is the procedure of analysing insights, discovering new paradigms and data and assuming future trends for a business in the existing marketplace. When the term data mining is generally used as the synonym for exploration of knowledge in databases, it is actually just one of the phases in the KDD procedure which is known as knowledge discovery in databases [7]. The major objective of KDD is to retain impactful and often unknown insights from a large set of data in a specific business.

Combining data mining and cyber security gives the permit for fixing up the features of cyber-attacks and develops attack detection processes within a business in a certain manner. Also, the data mining process is an effective process in a business and data can be protected for a particular form of security of a business. In several types of cases related to the cyber security of a business, the data mining can be used to mitigate the threats which can affect the privacy of a business and also, can harm the privacy protocols of a particular business. The process of data mining helps in malware identification in a particular case of a business. When constructing a specific type of software which regulates the security related works within a business, the developers implement the methods of data mining to develop the pace and quality of the "malware detection" also, to identify zero-day attacks as well. In the process of malware detection, there are three major types of strategies and the strategies are as follows- anomaly detection, detection of misuse and hybrid detection [8]. The anomaly detection includes the techniques of modelling of the normal behaviour of a system or network within the order to identify differences from normal action patterns.

However, the anomaly detection strategies can identify prior unknown threats and can be implemented for describing the signatures for misled detectors. Also, the "anomaly detection" can report even authorized actions in case it got diverted from the norms of, therefore, manufacturing the false positive alerts in a successive manner [9]. Also, another strategy of malware detection is misused detection. This particular strategy identifies the known attacks such as signature based attacks or threats in a certain manner. The following technique consists of lower rate of false positives however cannot detect zero days' attacks. Furthermore, another strategy of malware detection is known as hybrid approach. Hybrid approach generally gathers "anomaly and misuse detection" strategies to enhance the counts of traced intrusions while decaying the counts of false positives at a maximum rate.

Instead of constructing a usage of both legitimate programs and "malware programs" to make a classifier, it creates norms which have been regulated by the data mining algorithm in a certain manner. After the evaluation of the process, the anomaly detection phases of the systems search for differences from the "normal profile" and "misuse detection" from the part of the systems looks for the malware signatures in the code in a certain order. Furthermore, there are two major types of malware detection process which can detect any sorts of intrusion within the cyber security of a business and the two processes are as follows- extracting malware features and classifying and clustering the malwares [10]. At the initial phase of malware detection, the data mining algorithms extract or mitigate malware features from the records on "API calls", "n-grams binary strings", "program behaviour" and other events. A business owner can easily apply static, dynamic or hybrid examination to carry out the malware features from effectively junked files.

Furthermore, by implementing the techniques of ML, each specification algorithm builds a model that portrays both benign and malignant classes within a business program of a particular business. Also, the classifiers of training implementing such samples collection of files gives the permits to detect the recently released malware within a business program in an effective manner. There is another part of using data mining within the process of cyber security of a business and the process is known as the induction deception [11]. Threats can evaluate malware instructions with the help of the network, databases, web clients and regulating systems of a specific organization. By implementing data mining strategies or techniques within a business, a business owner can examine and audit the outcomes and fetch the unrecognized patterns. Therefore, a business owner can detect the unwanted intrusions, networks and systems scanning, denial of services and work penetration as well for a certain time span.

Furthermore, data mining attacks can be based on two effective types and the types are as follows- host related attacks and network based attacks or threats. In order to detect the host related threats, the cyber security software of a business is required to examine the features which have been extracted from the programs. By detecting network dependent threats, requires a resolvement to examine the networks related to traffic. Also, as with eh malware detection, a business owner can look for either anonymous behaviours or cases of misusing programs within a certain order. Intrusion detection systems have been generally dependent on the specification, clustering and relation norms techniques or strategies [12]. These strategies give a permit for extracting threats as features from a database, organizing the threats and flagging any new or upgrade records with the similar features. Among those features, some algorithms can be used by a business owner to implement it as a regression or decision tree.

A business owner can also add assumption abilities to the unwanted intrusion detection system in a certain manner to trace the intrusion within the business programs. Strategies such as classification and series of times examinations can evaluate the probability of a future intrusion within business programs. Also, by using artificial intelligence algorithms, helps to create easier detection or prior unrecognized and suspicious action which is associated with a particular business furthermore, the data mining helps to detect frauds within the cyber security of a business in a proactive form [13]. Data mining strategies that increase the usage of machine learning can be able to pick up several kinds of fraud, from economic fraud to telecommunication fraud and computing intrusion for a business. ML is specifically impactful for fraud detection because it can evaluate several tasks which are associated with a certain type of business.

The several types of deeds are as follows - scale to undertake within account within the count and difficulty of databases, learn to fetch and assume upgrade types of frauds and appropriately examine the portability of fraudulent actions. Additionally, a business owner can be able to implement both recognized and unrecognized ML algorithms to fetch fraud within a business. With the help of recognized learning, accessible details and records are differentiated as either "fraudulent" or "non-fraudulent" to business programs [14]. Later on, by using the differentiating model has been implemented for training a model to fetch possible fraud within the business organization which are related with a business from all aspects. The major challenge with this procedure is its fragile nature to fetch upgraded types of attacks which make the entire performance of a business at its lower cost.

On the other hand, for implementing unrecognized learning strategies, known fraud paradigms from untagged records make their own differentiation and feature introductions for fraudulent actions as soon as possible. Also, the unrecognized learning helps to fetch privacy protocols and security check-ins in data without implementing statistical examination. It also has the capability for examining and detecting news types of fraud within types of a business. There is another threat which can be mitigated by using data mining within a business and the threat is related to intelligence collection for a particular business [15]. Several pieces of proof about cyber security attacks are generally scattered over an organizational network and work accordingly with that organization in a certain manner. These records are generally implemented to construct training databases, construct data mining models and develop accuracy for each assumption in a certain manner. The threat management is related with the implementation of AI for discussion of the issues and benefits of AI to answer all the queries.

Data mining algorithms help to explore the corrupted and hidden data and change it within a constructed attack intelligence database. A business owner can be able to implement the norms of clustering, relating norms and briefing strategies to explore few types of intelligence to regulate the threats accordingly and intelligence are as follows-tactical, operational and strategic [16]. Data mining has been often implemented only for the initial phase among these three intelligences to explore and construct the data. Also, after this trivia, an expert from cyber security has to manually check the explored data and make decisions on how to act on it. However, a business owner can use data mining to develop frameworks which would be based on machine learning for collecting and processing data in a certain order.

Another benefit of data mining there in cyber security is to detect the insider threats and assume the attacks previously. Insider attacks are actions of authentic users that might be the cause to affect an organization from all aspects. Detecting those insider attacks actions is generally a strategic task because these activities often look the same to the ordinary user actions or the activities can be intentionally masked from the mechanism of attack detection from a certain manner. From the beginning of big data algorithms that can trace unwanted behaviour of machines and human users, they have been implemented vigorously to trace and assume insider attacks in a more suitable order. Similarly, to intrusion detection within a system, insider threat detection systems are dependent on tracing the traits of authentic and threatening activities [17]. There is a huge variation of machine learning based specifications and clustering algorithms, both of the viewed and none viewed that helps to trace insider attacks. Also, a business owner can provide training to "deep neural networks" on the basis of data mining norms and ethics to analyze cyber security logs and detect the possible insider actions in the real time world in all aspects of cyber security.

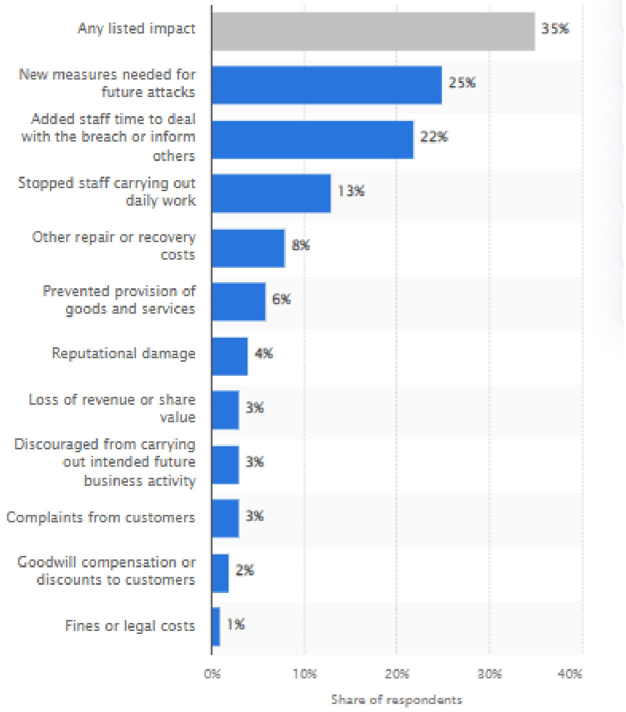

Figure 1: Impact of cyber security on the United Kingdom businesses in the year of 2022

By looking over a survey of the business of the United Kingdom in the years of October 2021 and January 2022, it has been found that 25 % of the reviewed businesses that have been experiencing data breaches, have faced the outcomes in new measures needed for further cyber attacks [18]. Above graph indicates that some businesses have faced "cyber security breach" or threat within the last twelve months. 22% of businesses which have been surveyed stated that staff time have been included to deal with the bread or inform others in a certain manner.

DISCUSSION

The following study is based on the cyber security and the implementation of data mining techniques within the cyber security of a business and the study has been served with proper sorts of insights related to cyber security and data mining. At the beginning of discussion, the concept of cyber security has been represented and after giving the conception on cyber security, the importance of cyber security has been discussed with elaboration. The importance of data mining in daily life has been discussed with sheer elaboration and later on after this following discussion, the importance of data mining techniques in a business have been mentioned in a certain manner. Furthermore, the impact of data mining techniques on keeping the confidential data in a particular order has been represented within an extended manner. Also, the detection of viruses by implementing the techniques of data mining has been shown in the following context and the virus detection is important to a business to prevent the attacks and threats of viruses in the business programs and decision making models as well.

Also, techniques related to data mining have been discussed which can predict and detect the malware in a business and also, it has been shown that how the malware can affect the entire flow of a business in an adverse manner to decrease the rate of profit in a business at a maximum rate. After the execution of the importance of data mining within a business, cyber security and its advantages in a business have been depicted within a certain manner. Also there are several kinds of techniques that are discussed which are related to data mining and can be used as the key player to detect the attacks related to malware, intrusion and several kinds of business oriented cyber malfunction can be regulated by the data mining techniques.

After the execution, several crucial sorts of implementation of data mining have been illustrated in the cyber security of a business. There are five major types of usages which can be evaluated by using data mining techniques in the cyber security of a business. As the prior usage, the malware detection has been evaluated and the malware can be detected by implementing data mining techniques in a successive manner. Also, there are several kinds of strategies have been depicted within the malware detection of a business and later on, the intrusion detection have been measured in a certain manner of using the data mining techniques within a business program. Later on, there are several types of usages that have been discussed in an extensive manner.

CONCLUSION

The following study is based on the concept of cyber security of a business and the impact of data mining techniques to strengthen the cyber security in successive manner and throughout the study, the focus has been made over the importance of data mining within the cyber security of a business. In recent days, fraudulent activities are encountered by businesses; however, firms need to focus on enhancing cyber security by prioritising data mining techniques. At the beginning of the study, an introductory segment has been served which consist of a decent amount of information regarding cyber security and data mining techniques in a business. In this section, a brief account has been made over the topic of cyber security and the concept of data mining within an extensive manner. Later on in the following part, the importance has been extended by gathering proper insights related to a business. After the execution of the following part, the materials and methods have been selected and evaluated to bring betterment in the execution process of the following subject matter. There are some materials and methods that have been selected yet all are immensely crucial to give the study a proper execution in a certain manner.

All the data which has been collected are secondary data and the data have been collected by following qualitative methods. In this following study, the inductive approach and cross sectional research design have been selected and implemented in a certain manner to give this study a proper. Later on, several kinds of usages of data mining techniques have been depicted to strengthen cyber security of a business and plenty of techniques have been shown to justify the subject matter in a successive manner. The techniques generally have been used to detect all sorts of malware and threats within a business programs in a proper order to mitigate the rate of malwares within a business program in a proper manner within the following study.

REFERENCES

[1] Sarker, Iqbal H., et al. "Cybersecurity data science: an overview from machine learning perspective." *Journal of Big data* 7.1 (2020): 1-29.

[2] Corallo, Angelo, Mariangela Lazoi, and Marianna Lezzi. "Cybersecurity in the context of industry 4.0: A structured classification of critical assets and business impacts." *Computers in industry* 114 (2020): 103165.

[3] Li, Ling, et al. "Investigating the impact of cybersecurity policy awareness on employees’ cybersecurity behavior." *International Journal of Information Management* 45 (2019): 13-24.

[4] Alloghani, Mohamed, et al. "Implementation of machine learning and data mining to improve cybersecurity and limit vulnerabilities to cyber attacks." *Nature-Inspired Computation in Data Mining and Machine Learning*. Springer, Cham, 2020. 47-76.

[5] Gupta, Brij B., and Quan Z. Sheng, eds. *Machine learning for computer and cyber security: principle, algorithms, and practices*. CRC Press, 2019.

[6] Rekha, Gillala, et al. "Intrusion detection in cyber security: role of machine learning and data mining in cyber security." *Advances in Science, Technology and Engineering Systems Journal* 5.3 (2020): 72-81.

[7] Salloum, Said A., et al. "Machine learning and deep learning techniques for cybersecurity: a review." *The International Conference on Artificial Intelligence and Computer Vision*. Springer, Cham, 2020.

[8] Thach, Nguyen Ngoc, et al. "technology quality management of the industry 4.0 and cybersecurity risk management on current banking activities in emerging markets-the case in Vietnam." *International Journal for Quality Research* 15.3 (2021): 845.

[9] Aiyanyo, Imatitikua D., Hamman Samuel, and Heuiseok Lim. "A systematic review of defensive and offensive cybersecurity with machine learning." *Applied Sciences* 10.17 (2020): 5811.

[10] Sarker, Iqbal H., et al. "Intrudtree: a machine learning based cyber security intrusion detection model." *Symmetry* 12.5 (2020): 754.

[11] Paul Joseph, D., and Jasmine Norman. "An analysis of digital forensics in cyber security." *First international conference on artificial intelligence and cognitive computing*. Springer, Singapore, 2019.

[12] Li, Kuan-Ching, Xiaofeng Chen, and Willy Susilo, eds. *Advances in Cyber Security: Principles, Techniques, and Applications*. New York, NY, USA: Springer, 2019.

[13] Ferrag, Mohamed Amine, Messaoud Babaghayou, and Mehmet Akif Yazici. "Cyber security for fog-based smart grid SCADA systems: Solutions and challenges." *Journal of Information Security and Applications* 52 (2020): 102500.

[14] Sun, Ruxia, et al. "Strategies for data stream mining method applied in anomaly detection." *Cluster Computing* 22.2 (2019): 399-408.

[15] Sreedevi, A. G., et al. "Application of cognitive computing in healthcare, cybersecurity, big data and IoT: A literature review." *Information Processing & Management* 59.2 (2022): 102888.

[16] Kalinin, Maxim, Vasiliy Krundyshev, and Peter Zegzhda. "Cybersecurity risk assessment in smart city infrastructures." *Machines* 9.4 (2021): 78.

[17] Zeadally, Sherali, et al. "Harnessing artificial intelligence capabilities to improve cybersecurity." *Ieee Access* 8 (2020): 23817-23837.

[18] Statista Research Department, “Have the breaches or attacks experienced in the last 12 months impacted your organization in any of the following ways, or not?” *Statista*, 10th January, 2023. https://www.statista.com/statistics/586747/impact-of-cyber-security-breaches-experienced-by-united-kingdom-uk-businesses/