Cyber Security Event Detection Using Machine Learning Technique

Salman Muneer[1], Muhammad Bux Alvi[2], AminaFarrakh[3]

[1]University of Central Punjab, Lahore, Pakistan.
[2]The Islamia University of Bahawalpur, Bahawalpur, Pakistan
[3]School of Management Sciences, Comsats University, Islamabad, Pakistan
Corresponding Author: Salman Muneer: salman.muneer@ucp.edu.pk

Abstract- Artificial Intelligence and Machine Learning techniques have become crucial components in the field of cybersecurity. The proposed model you mentioned can help to enhance the overall security of a system by detecting and preventing malicious attacks in real-time. With the help of advanced algorithms, machine learning models can analyze large amounts of data, identify patterns and anomalies, and take appropriate action to prevent a security breach. This leads to improved detection rates and reduced false positive rates, which results in a more effective defense against cyber threats.
Additionally, the model can be used to develop new security frameworks for companies and organizations. These frameworks can include network security, data protection, device security, and identity management, among others. This research presents a cyber security-based model helps to ensure that all critical assets are protected from cyber-attacks and that sensitive information is not leaked or stolen. The integration of AI and ML techniques into cybersecurity systems has the potential to significantly improve the overall security posture of an organization and help protect against the growing threat of cyber-attacks.

Keywords: Event Detection, cyber security, machine learning

1 Introduction

Cybersecurity is preserving cyber threats to internet-connected systems, including data, software, and hardware. In order to prevent unauthorized access to data centres and other computerized systems, people and businesses use this technique. There are many types of cyber security. However, one of them is provided by Kaspersky Labs: Digital protection is the method of forestalling destructive attacks on PCs, servers, cell phones, electronic frameworks, organizations, and information. It is frequently alluded to as electronic data security or data innovation security. The phrase has many applications and can describe everything from end-user training to disaster recovery [1].

Individual, legislative, and business information should be shielded from abuse or control by outsiders. In order to do this, cybersecurity must focus on three essential tasks: (a) protecting hardware and software packages, as well as the information they include; (b) ensuring the protection's government or value against various threats; and (c) putting these tasks into practice and improving them. Although digital assaults utilize no actual weapons, they are the riskiest and most hurtful weapons that might cause disclosure of the highest arranged data of government associations through surveillance or delicate individual data through to phishing. As per network safety specialists, simply in 2017, digital assaults could have caused US$5 billion worth of harm and will fill from now on. For instance, the damage might hit US$6 trillion yearly by 2021. A few counter-measures against digital assaults have been presented for many years, mostly known as interruption discovery frameworks (IDSs). Lately, computational knowledge methods, including AI (ML), have been used to guarantee network safety. Regardless of wonderful advancements in the utilization of computational knowledge methods and ensuing expansion in exhibitions, heartiness against digital assaults and experiences of malignant examples and assaults, computational knowledge in online protection actually needs to progress significantly, other than defeating many difficulties, for example, zero-day assaults. Besides, there is likewise a developing worry about the security and weaknesses of ML strategies against assaults [2].

In order to enhance the models, it is necessary to consider all of the capabilities of machine learning approaches, taking into account numerous variables like calculation time, actualization ability, and difficulty. The importance may alteration depending on the application. Beyond the error rate, other

performance metrics are taken into account because they have demonstrated several benefits. The limit of programmed classifiers to accurately recognize malware has been tried, handling the bogus up-sides cases and successes utilizing classifiers in light of perception. Artificial intelligence strategies have created danger location frameworks, utilizing Bayesian regularized brain organizations. Examples of classifiers include naive Bayes, Bayesian classifiers, support vector machines (SVM), classifiers based on brain networks, and self-association guidelines and references inside. Goseva-Popstojanova and associates presume that AI procedures, for example, SVM and choice trees, can effectively recognize assault Web meetings. Fluffy rationale and brain networks have been effectively consolidated for malware discovery, concentrating on the main API calls. Phishing assaults are specific wrongdoing that gets individual data from clients through deceitful sites and is the most widely recognized technique for fraud. AI methods have been utilized to distinguish the origin of the phishing assault as well as to recognize phishing messages contrasting several AI methods [3].

In order to identify threats in cyberspace, numerous techniques and procedures have been created. Although viruses, whose evolution and development are typically quicker than malware detection software development, can bypass their procedures, commercial software (antivirus) can be utilized in the case of malware with good results. Due to the threats' quick evolution, learning approaches, such as various machine learning algorithms, have been adopted to identify new viruses. This research can measure the cutting edge of AI strategies utilized in network safety by involving the Scopus data set as the essential wellspring of this writing audit. In June 2015, the pursuit "AI AND spam" returned 473 outcomes; "AI AND malware," 326 outcomes; lastly, "machine learning AND phishing," 94 results. The following sections will concentrate on these three dangers since, in light of these numbers, they are thought to be the most significant (or the most researched) [4].

2 Related work

Numerous studies have been conducted in cybersecurity to identify cyber-attacks, cyber-irregularities, and intrusions. Anomaly-based intrusion detection systems (AIDS) and signature-based intrusion detection systems (SIDS) are well known in the cyber business for identifying and stopping cyberattacks. SIDS is based on recognized attack signatures. Over SIDS, AIDS has the advantage of being able to spot invisible risks, such as the distinction between undiscovered or zero-day assaults. Although association analysis is widely used in machine learning to create rule-based intelligent systems, its effectiveness in identifying abnormalities or cyberattacks may be limited by its repetitious creation and difficulty with increasing safety elements. Due to their intuitive understanding abilities after protecting information, machine learning classification models for security modelling are the main focus of this work to attain our goal [5]. A few investigations have used a strategic recurrence rule to identify vengeful traffic and interruptions. The KNN, a case-centred knowledge calculation, is one more typical technique for AI where the characterization a not set in stone by that information point's k-closest neighbours. Vishwakarma et al. use the KNN arrangement strategy in the examinations with the end goal of interruption identification frameworks. Creators in think about brain classifier, as well as in think wavelet change for oddity discovery especially DoS assaults. Countless examinations in the area of network protection, for example, Relan et al. employ the DT arrangement strategy, as do Rai et al., Ingre et al., Malik et al., Puthran et al., Moon et al., Balogun et al., and Sangkatsanee et al. in their examinations to fabricate interruption identification frameworks. To recognize oddities and address loT network safety dangers in shrewd RF picking up comprising of different choice plants in the paired arrangement model. Mazini et al. use the AdaBoost approach to highlight determination but build an abnormality web-centred interruption recognition framework in their work. An AI protection standard for recognizing peculiarities has been introduced, which is viable as far as pre-style exactness and diminishing the component aspects because of the choice decision tree approach with highlight determination. As of late, an AI-based botnet assault recognition system with consecutive location design has been introduced where ANN, DT, and NB grouping strategies are utilized. Hasan et al. perform assault location examination in IoT locales to foster a shrewd, solid, IoT-based foundation. Albeit a few AI strategies, like SVM, DT, RF, LR, and ANN, are utilized, the examination is restricted to a few security highlights for identifying various kinds of assaults.

Also, the varieties in the meaning of the security highlights, which could be a significant part while building a viable security model utilizing AI methods, are not tended. In reality, the network protection issues may be engaged with countless security highlights, and the viability of a learning-based security model might differ depending on the meaning of the related security highlights and the information qualities. Different sorts of AI strategies and their materialness in cybersecurity have been discussed in Sarker et al. An itemized experimental examination is expected to go with an intelligent choice in the vicinity. Dissimilar to the above, this paper presents "CyberLearning", AI-based cyber security displaying with related highlight determination as per their importance in demonstrating, and a thorough experimental examination on the viability of different AI-based security models. While building the security models, we consider a double order model for recognizing oddities and a multi-class characterization model for identifying multi-assaults with regards to network safety. In order to give a complete view to the perusers nearby, we likewise sum up the most pertinent AI-based security models inside the extent of our review for an unmistakable comprehension for the peruses [6].

Most of the approaches have been used while employing and constructing several smart as well as intelligent frameworks like machine learning algorithms [7-9], Particle Swarm Optimization [10], Fusion based approaches [11], cloud computing [12], transfer learning [13], MapReduce [14] and data security and privacy systems [15] that may provide assistance in designing emerging solutions for the rising challenges in designing smart cloud-based monitoring management systems.

3 Proposed Methodology

The growth of artificial intelligence (AI) methods has resulted in a significant increase in the development of learning-based approaches for detecting numerical attacks. These approaches have produced outstanding results in many evaluations and have become an essential tool in protecting IT systems against threats and malicious behaviors in networks. However, the constantly evolving nature of digital attacks makes it challenging to secure networks effectively. The need for robust defenses and protection measures has become imperative due to the growing number of government intrusions and malicious activities. One of the critical challenges in network security is to design a digital threat detection process that is automated and effective. To address this issue, this research proposes a cyber-event detection model that can provide efficient results in predicting cyber events. The proposed model integrates advanced machine learning techniques and real-time network data analysis to detect and prevent cyber-attacks. The model is designed to detect both known and unknown cyber threats, providing a proactive approach to network security. The use of real-time data analysis and machine learning algorithms allows the model to constantly adapt and improve its accuracy, making it a highly effective solution for cyber-event detection.

Figure 1 presents a visual representation of the proposed cyber-event detection model. The model consists of multiple components, including data collection, data preprocessing, feature extraction, feature selection, and model training. Data collection is performed using various sources, including network logs, system logs, and intrusion detection systems. The collected data is preprocessed to remove any irrelevant or redundant information and to format it in a manner suitable for analysis. Feature extraction and selection are performed to identify the most important features in the data that can effectively predict cyber events. Finally, the model is trained using the selected features, and its performance is evaluated using various metrics. The proposed cyber-event detection model provides a promising solution for detecting and preventing cyber-attacks. Its ability to detect both known and unknown threats and its continuous adaptability make it an attractive solution for organizations looking to secure their networks.

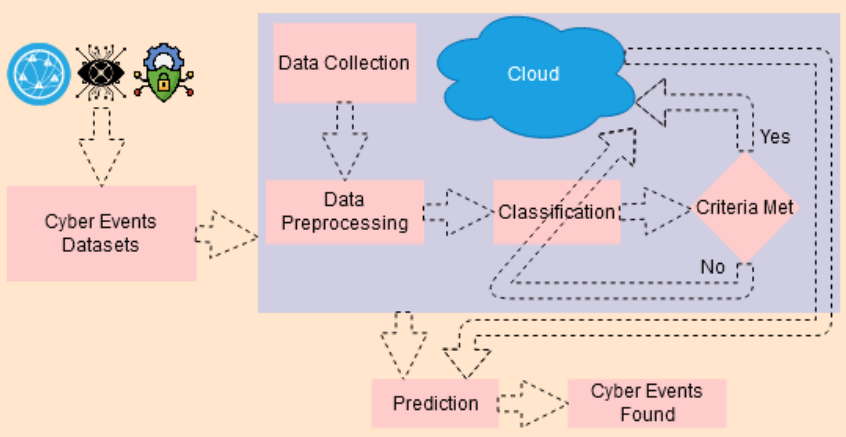

Figure 1: Proposed energy management model

As depicted in Figure 1, the proposed cyber-event detection model involves several stages of data processing and analysis. The first stage is data collection, where relevant information about cyber events is gathered through various input parameters and stored in a database for later analysis.

The collected data is then subjected to data preprocessing, where any noisy or irrelevant information is removed to prepare the data for the classification stage. During the classification stage, machine learning-based approaches are applied to diagnose the presence of cyber events and to assess whether the learning criteria have been met. If the criteria are not met, the classification process is retrained, and if they are met, the trained outcome is stored on a cloud-based platform.

Once the trained outcome is stored on the cloud, it can be imported for predicting purposes. The model uses the imported data to determine the likelihood of a cyber event occurring, and if a cyber event is detected, a message is displayed indicating its discovery. The proposed model provides a comprehensive and automated approach to detecting and preventing cyber events, leveraging advanced data analysis and machine learning techniques to deliver accurate and reliable results.

4 Limitations and Future Directions

Today's technology-driven world, it is crucial to identify emerging security threats from open data sources as a part of ensuring the security of deployed software and systems. However, current security systems face numerous limitations, such as data breaches caused by remote work, ransomware attacks, unpatched vulnerabilities, Bring Your Own Device (BYOD) threats, and lack of backup plans. To address these challenges, a cyber-security-based model has been proposed that leverages machine learning techniques. The proposed model aims to provide a comprehensive solution to these security issues and will play a crucial role in shaping the future of cybersecurity. As companies strive to secure their networks, data, devices, and identities, the adoption of security frameworks such as zero trust will become increasingly important. Zero trust will help companies secure internal information systems and data stored in the cloud, thus providing an extra layer of protection against cyber threats.

5 Conclusion

The use of machine learning in cyber security event detection is a promising development in the field of information security. By leveraging advanced algorithms and techniques, such as supervised and unsupervised learning, machine learning models can identify patterns, anomalies and correlations in vast amounts of data that would be difficult to detect by human analysts alone. Furthermore, the ability of machine learning models to continuously learn and adapt to new data and changes in the threat landscape provides organizations with the ability to stay ahead of evolving cyber security threats. However, it's crucial to note that the deployment of machine learning in cyber security event detection must be done with caution, as the potential for false positive or negative detections can have serious consequences for an organization's security posture. As such, a robust evaluation and validation process should be in place to ensure the quality of the models and their effectiveness in detecting real-world threats.

6 References

[1] Von Solms, R., Van Niekerk, J., 2013. From information security to cyber security. Comput. Secur. 38, 97–102.

[2] Siddique, K., Akhtar, Z., Khan, M.A., Jung, Y.H., Kim, Y., 2018. Developing an intrusion detection framework for high-speed big data networks: A comprehensive approach. KSII Trans. Internet Inf. Syst. 12, 4021–4037.

[3] Benaddi, H., Ibrahimi, K., 2020. A Review: Collaborative Intrusion Detection for IoT integrating the Blockchain technologies. Proc. - 2020 Int. Conf. Wirel. Networks Mob. Commun. WINCOM 2020.

[4] Geetha, R., Thilagam, T., 2021. A Review on the Effectiveness of Machine Learning and Deep Learning Algorithms for Cyber Security. Arch. Comput. Methods Eng. 28, 2861–2879.

[5] Sarker, I.H., 2019. Context-aware rule learning from smartphone data: survey, challenges and future directions. J. Big Data 6, 1–25.

[6] Bapat, R., Mandya, A., Liu, X., Abraham, B., Brown, D.E., Kang, H., Veeraraghavan, M., 2018. Identifying malicious botnet traffic using logistic regression. 2018 Syst. Inf. Eng. Des. Symp. SIEDS 2018 266–271.

[7] Aslam, M.S., Ghazal, T.M., Fatima, A., Said, R.A., Abbas, S., Khan, M.A., Siddiqui, S.Y., Ahmad, M., 2021. Energy-efficiency model for residential buildings using supervised machine learning algorithm. Intell. Autom. Soft Comput. 30, 881–888.

[8] Ghazal, T.M., Noreen, S., Said, R.A., Khan, M.A., Siddiqui, S.Y., Abbas, S., Aftab, S., Ahmad, M., 2022. Energy demand forecasting using fused machine learning approaches. Intell. Autom. Soft Comput. 31, 539–553.

[9] Khan, M.F., Ghazal, T.M., Said, R.A., Fatima, A., Abbas, S., Khan, M. A., Issa, G.F., Ahmad, M., Khan, Muhammad Adnan, 2021. An iomt-enabled smart healthcare model to monitor elderly people using machine learning technique. Comput. Intell. Neurosci. 2021.

[10] Asadullah, M., Khan, M.A., Abbas, S., Alyas, T., Saleem, M.A., Fatima, A., 2020. Blind channel and data estimation using fuzzy logic empowered cognitive and social information-based particle swarm optimization (PSO). Int. J. Comput. Intell. Syst. 13, 400–408.

[11] Ihnaini, B., Khan, M. A., Khan, T.A., Abbas, S., Daoud, M.S., Ahmad, M., Khan, Muhammad Adnan, 2021. A Smart Healthcare Recommendation System for Multidisciplinary Diabetes Patients with Data Fusion Based on Deep Ensemble Learning. Comput. Intell. Neurosci. 2021.

[12] Gai, K., Guo, J., Zhu, L., Yu, S., 2020. Blockchain Meets Cloud Computing: A Survey. IEEE Commun. Surv. Tutorials 22, 2009–2030.

[13] Muneer S, Akhtar A, Qamar HU. Revolutionizing Smart Cities through Transfer Learning: A Comprehensive Review. International Journal of Computational and Innovative Sciences. 2023 Mar 30;1(1):40-4.

[14] Asif, M., Abbas, S., Khan, M. A., Fatima, A., Khan, Muhammad Adnan, Lee, S.W., 2021. MapReduce based intelligent model for intrusion detection using machine learning technique. J. King Saud Univ. - Comput. Inf. Sci.

[15] Aslam, M., Khan Abbasi, M.A., Khalid, T., Shan, R.U., Ullah, S., Ahmad, T., Saeed, S., Alabbad, D.A. and Ahmad, R., (2022). Getting Smarter about Smart Cities: Improving Data Security and Privacy through Compliance. Sensors, 22(23), p.9338.