Automated Threat Correlation Using Machine Learning: A Framework for Enhanced Cybersecurity

FEYISAYO OGUNMADE
University of Louisville

Abstract- ***As cyber threats grow increasingly sophisticated, the demand for agile, accurate, and automated methods to detect and respond to attacks has become imperative. This paper proposes a novel machine learning-driven framework for automated threat correlation, aimed at enhancing real-time threat detection and minimizing the manual oversight often required in traditional cybersecurity measures. By leveraging advanced algorithms, such as k-means clustering for event grouping, neural networks for pattern recognition, and the Apriori algorithm for association rule mining, the framework is designed to correlate threats from diverse data sources, including network traffic logs and threat intelligence feeds. This integration of machine learning models enhances detection accuracy, reduces false positives, and accelerates response times, significantly improving resource allocation for cybersecurity teams. The proposed framework also addresses key challenges in data preprocessing, model selection, and privacy compliance, demonstrating its potential for scalability and adaptability to various threats. Comparative analysis with prior approaches highlights the framework's efficiency in reducing detection latency and improving resilience against multi-stage cyberattacks. This work concludes with recommendations for future enhancements, such as incorporating deep learning models and expanding data sources, to further refine the framework's capabilities. The proposed machine learning-based approach for automated threat correlation represents a critical advancement in cybersecurity, providing organizations with an adaptive, resilient, and scalable solution.***

Indexed Terms- ***Cybersecurity, Machine Learning, Threat Correlation, Automated Detection, Clustering, Neural Networks, Real-Time Threat Detection, Data Preprocessing, False Positives, Cyber Threat Intelligence***

I. INTRODUCTION

The world has shifted towards digital, cybersecurity threats are escalating in complexity and frequency, posing a critical risk to national security, economic stability, and democratic institutions. The prevalence of cyber-attacks targeting U.S. businesses and governmental institutions has been escalating steadily, posing significant financial and operational challenges. According to an FBI report, these attacks resulted in financial losses exceeding $12.3 billion in 2023 alone. This alarming trend underscores the urgent need for organizations to strengthen their cybersecurity frameworks, particularly in developing advanced threat detection and response mechanisms. Effective strategies are crucial for preventing large-scale data breaches and minimizing the associated risks and damages (Statista, 2024; Aminu et al., 2024). Traditional threat detection methods often struggle to keep pace with the volume and sophistication of contemporary attacks, stating the urgent need for faster, more accurate threat correlation capabilities (Thapliyal et. al., 2024).

The cybersecurity space faces new challenges as modern cyber threats become more advanced and occur more frequently (Salem et al., 2024). With the proliferation of cyber threats, ranging from ransomware, DDoS, man-in-the-middle, and phishing attacks to advanced persistent threats (APTs) which traditional security can no longer prevent, many current systems cannot sift through vast data logs quickly enough to identify attacks before significant damage occurs (Aslan et al., 2023). The IBM 2024 report reveals that 67% of organizations now use security AI and automation, with a 98-day faster incident response time and a reduced breach lifecycle of 258 days. Additionally, internal detection has risen to 42%, saving companies nearly $1 million and shortening breach lifecycles by 61 days (IBM, 2024). Effective cybersecurity hinges on the ability to detect patterns that signal potential threats across diverse data

sources, such as network traffic, endpoint activities, and user behavior logs (Wasyihun et al., 2024). Threat correlation involves using correlation engines within SIEM systems to track and analyze behavior patterns across various assets, which generate both disparate and related events, to identify potential threats (Cybersainik, 2021). Without effective threat correlation, organizations may miss early attack indicators, as traditional IDS systems often struggle to link alerts from different sources, allowing sophisticated, distributed attacks to go undetected and cause greater damage (Haas et al., 2019).

The objective of this article is to propose a machine learning (ML)--driven framework for automating threat correlation. By leveraging advanced analytics, the proposed framework aims to enhance the speed and accuracy of threat detection, offering a powerful tool for cybersecurity professionals in their fight against increasingly sophisticated threats. This framework would utilize machine learning algorithms to automatically correlate data from multiple sources, flagging potential threats and allowing security teams to respond swiftly. Given the advancements in artificial intelligence and data science, implementing an ML-based threat correlation system presents a promising solution for the cybersecurity challenges facing U.S. organizations today.

II. LITERATURE REVIEW

Threat correlation in cybersecurity has evolved as a critical approach for identifying and mitigating sophisticated cyber threats. Traditional methods largely rely on manual threat analysis, in which cybersecurity analysts review security logs and attempt to connect disparate data points to detect threats, which can be slow and error-prone (Maosa et al., 2024). However, with the exponential increase in both the volume and complexity of attacks, manual correlation is increasingly inadequate.

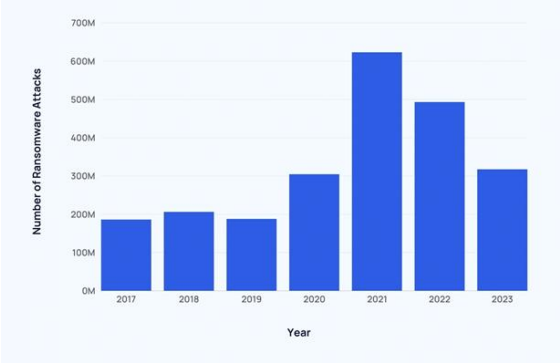

Fig 1: Worldwide Ransom Attacks
Source: Exploding Topics

The evolving role of machine learning in making threat correlation both actionable and scalable is better suited to handle the high volume and variety of modern cybersecurity data alerts more effectively (Gomes et al., 2022). An ideal Intrusion Detection System (IDS) would detect all attacks with perfect accuracy (100% detection rate) while generating no false positives (0% false positive rate); however, achieving this balance is exceptionally challenging in practice (Norsyafawati, 2011). In response, several researchers have emphasized the need for more adaptable, data-driven correlation methods that can analyze large datasets and dynamically identify patterns indicative of emerging threats.

Machine Learning in Cybersecurity

Machine learning (ML) has shown significant potential in cybersecurity, particularly in enhancing threat detection through clustering (unsupervised), classification (supervised), and association (unsupervised) algorithms. Clustering algorithms such as k-means and density-based clustering non-parametric algorithms (DBSCAN) are frequently used to group related events and identify anomalous patterns, which can signal potential security breaches (Arévalo et al., 2022). A research by Sameera et al. (2023) highlighted a clustering framework NSL-KDD that could detect known and Zero-Day attack patterns with 78% accuracy, offering a substantial advantage over traditional methods. In the context of machine learning for cybersecurity, classification algorithms such as decision tree algorithms have demonstrated remarkable efficiency, achieving 98.89% accuracy in just 0.098 seconds, as shown by Chaganti and Tadi (2022). Similarly, neural networks, as explored in

Dalal et al.'s research, achieved an accuracy rate of at least 99.09%, with a false negative rate capped at 6.75%. The ability of these algorithms to reduce false positive and false negative rates by over 65% enhances their suitability for real-time threat detection, enabling faster, more accurate identification and response to cyber threats. This progress highlights machine learning's potential to address critical challenges in automated threat correlation effectively. According to an empirical study by Pavithra & Selvakumara Samy (2022), in an experiment on an adware dataset, the Random Forest classifier outperformed Naive Bayes and SVM across two test scenarios. It achieved a remarkable accuracy of 0.9947 in adware classification and demonstrated lower time complexity on larger datasets compared to the other techniques, emphasizing their utility in proactive threat detection.
Association algorithms, meanwhile, focus on detecting co-occurrences among events, which is particularly useful in threat correlation. These algorithms can link multiple indicators, such as failed login attempts and unusual IP addresses, to flag potential threats. Research by Dol Aher, Sunita & Lobo. (2012) suggests that when association algorithms are integrated with clustering and classification, they offer a comprehensive approach to detecting complex, multi-stage attacks. The Apriori algorithm is employed to mine frequent item set properties, utilizing an iterative approach called level-wise search, significantly reducing the response time and enhancing detection rates, making it a better alternative to static rule-based systems (Kumar Mukesh, 2012).

Gaps in Current Research

Despite the advancements, significant gaps persist in automated threat correlation methods. Traditional rule-based approaches and even some machine learning methods still face limitations in detecting complex, multi-layered attacks that unfold over extended periods. Sumathi et al. (2023) indicate that rule-based systems struggle to capture the evolving nature of threats and are often reactive rather than proactive, detecting threats only after they have infiltrated a network. Furthermore, many machine learning models rely on labeled datasets to train classification algorithms, yet acquiring labeled data for new and emerging threats is challenging, leading to potential blind spots in the detection process (Mustafa, 2023; Ahsan, 2022).

Another gap in current research is the lack of real-time adaptability in most existing threat correlation models. While clustering and classification algorithms are highly effective in correlating threats across datasets, they may fail to adapt to new attack strategies due to their dependence on pre-existing data patterns. Kaspersky Labs (2022) has reported on this limitation, pointing out that many machine learning-driven security systems are vulnerable to adversarial attacks, especially in nonstationary environments, where threat actors modify their tactics to evade detection (Macas et al., 2024). These gaps emphasize the need for more adaptive and resilient models, potentially through reinforcement learning, which could enable the correlation system to learn and respond to new threats autonomously.

III. PROPOSED FRAMEWORK FOR AUTOMATED THREAT CORRELATION

Framework Overview

This framework with a focus on machine learning integration, offers an efficient and scalable solution to correlate threats across multiple sources in real time. By using advanced algorithms and a layered approach to data management and analysis, this framework is designed to address critical challenges in traditional threat correlation methods, such as high false positive rates and lengthy detection times. Unlike purely rule-based systems or manual threat correlation approaches, this framework’s machine learning-driven capabilities enable adaptive, rapid identification of evolving cyber threats.

Components of the Framework

Data Collection: Data collection is the foundational step, with a focus on capturing large amounts of information from diverse sources, including system logs, network traffic, and threat intelligence feeds. Security Information and Event Management (SIEM) systems, combining various data points, allowing enhanced visibility and situational awareness, essential for proactive threat detection. By integrating with established SIEM tools, the proposed framework ensures that data is comprehensive and capable of

providing multi-layered insights into potential threats (Fakiha, 2020).

Data Preprocessing: Effective threat correlation requires high-quality data, making preprocessing a critical phase. This step includes processes such as data cleaning, normalization, and feature extraction to ensure that raw data is optimized for machine learning models. Preprocessing reduces noise and standardizes information, creating a reliable dataset that can significantly improve detection accuracy. Studies have demonstrated the effectiveness and applicability of preprocessing techniques like data cleaning, transformation, normalization, and feature selection (Dhongade et al., 2024).

Model Selection: The choice of machine learning models directly impacts the framework's ability to detect and correlate threats. A Multi-stage Threat Correlation framework of Machine Learning, similar to Sarker et al. (2020) and Liu et al. (2021), this paper proposes a hybrid, Multi-stage framework for enhanced cybersecurity. Clustering algorithms are utilized to group related events, thereby isolating abnormal activity patterns. Association algorithms are applied to identify relationships within data points, detecting hidden patterns indicative of potential threats. Research by Chaganti and Tadi (2022) shows that decision trees can achieve high accuracy rates (98.89%) with minimal latency, while neural networks achieve even higher accuracy with lower false negatives, making them particularly suited for threat detection and correlation. A strategic integration of this clustering algorithm like k-means, with a combination of association algorithms like the Apriori association algorithm, using a neural network as a primary layer, will provide a cybersecurity system.

Threat Correlation: The framework's core lies in its ability to correlate diverse data inputs to identify complex, multi-stage cyberattacks. By correlating events across multiple sources, the model can spot coordinated attacks that may otherwise go unnoticed. Multi-stage attacks often involve multiple, seemingly unrelated events that, when analyzed together, reveal a larger threat pattern. Utilizing machine learning to automate this process allows for faster, more accurate identification of such patterns than traditional methods, which rely on manual intervention and rule-based approaches.

IV. CASE STUDY

Stellar Cyber's Open XDR Platform

Stellar Cyber implemented an AI-driven incident correlation solution using its Open XDR Kill Chain, fully compatible with the MITRE ATT&CK framework, designed to transform raw alerts into comprehensive incidents. This system combines machine learning and GraphML algorithms to group related alerts, dramatically improving detection accuracy and reducing analyst workload. With this approach, the platform reduced the mean time to detection (MTTD) by approximately 8 times and the mean time to respond (MTTR) by 20 times, ensuring improvements in both the speed and accuracy of threat response. This platform also minimizes alert fatigue by consolidating similar alerts, enabling faster, more precise detection and streamlined incident management. The technology's ability to focus analysts on fewer, high-fidelity alerts has significantly enhanced response efficiency for organizations using it, providing a strong example of how automated correlation can speed threat detection and response times effectively (Stellar Cyber, 2021).

Case Study 2; Correlation of Advanced Persistent Threats (APT) Using Behavioral Analysis

A study conducted by Li et al. (2024) employed a unique method of correlating Advanced Persistent Threat (APT) groups by analyzing behavioral patterns and using rough set theory. This approach allowed for the automatic identification of related behaviors across APTs, which improved the detection accuracy to over 90%. By focusing on common behavior patterns among APTs, this framework helped organizations prioritize resources and respond more efficiently while reducing the false-positive rates often associated with traditional rule-based threat correlation. The study demonstrated that automating the correlation of behaviors among similar threat actors could streamline resource allocation and enhance response efficacy, especially in complex threat environments where similar attacks occur across varied vectors (Li et al., 2024).

These case studies reflect the successful implementation of automated threat correlation. First, using AI-driven incident grouping, such as that in Stellar Cyber's system, can significantly reduce analyst fatigue and enhance response times by focusing attention on high-priority incidents. Second, using behavioral pattern analysis, as in the APT correlation study, shows that accurate threat correlation can streamline resource allocation and reduce the likelihood of false positives, helping organizations manage threats more effectively and efficiently. Both cases stated how automated threat correlation tools can offer substantial improvements in cybersecurity response metrics and resource management.

V. BENEFITS OF AUTOMATED THREAT CORRELATION IN CYBERSECURITY

Automated threat correlation in cybersecurity offers substantial benefits, including reduced response times, improved detection accuracy, and optimized resource allocation. Unlike traditional, manual methods that rely on human analysts to detect threats, automated systems can analyze data from multiple sources in real-time or near-real-time, largely improving the speed of threat detection and response. AI-driven threat detection that uses ML can identify, analyze, and improve response times allowing teams to address threats before they escalate (Potter & Lucas 2024). Accuracy is also significantly enhanced, as machine learning algorithms like neural networks and decision trees have shown detection rates exceeding 98% including ConXNet achieving more than 97% accuracy and 98% precision, minimizing both false positives and false negatives that can overwhelm human analysts and lead to critical oversights (Azeem et al., 2023; Chaganti & Tadi, 2022; Dalal et al., 2023). By automating repetitive tasks, these systems free up cybersecurity teams to focus on high-priority issues, making resource allocation more efficient and reducing the overall operational burden. By enhancing productivity and lowering costs, teams can effectively manage cybersecurity threats without scaling human resources to the same degree (Tonhausera & Ristveja, 2023).

VI. CHALLENGES AND CONSIDERATIONS

Automated threat correlation in cybersecurity, while effective, faces key challenges related to data quality, model scalability, and privacy compliance. Data quality and preprocessing present a significant challenge, as data must often be gathered from diverse sources, each with unique formats, protocols, and standards (Cai et al., 2015). Inconsistent or incomplete data can undermine model accuracy, requiring extensive preprocessing steps such as cleaning, normalization, and feature extraction to ensure that the data used is accurate and consistent (Maharana et al., 2022; Naseem, 2024). Additionally, large-scale datasets introduce scalability challenges, as machine learning models may require continuous tuning to handle growing volumes of data without compromising accuracy or performance (Sharma Vinod, 2022). Studies by Tufail et al. (2023) suggested that advanced models like neural networks and clustering algorithms may struggle with real-time processing as datasets increase, highlighting the need for scalable infrastructure and ongoing model optimization. Privacy and compliance are also critical, particularly when handling sensitive information that could expose organizations to regulatory scrutiny. Automated threat correlation must adhere to data privacy laws such as the General Data Protection Regulation (GDPR) or the California Consumer Privacy Act (CCPA) among others making it essential for frameworks to incorporate privacy-by-design principles. Balancing these technical and regulatory considerations is essential to developing a resilient, legally compliant threat correlation system that aligns with modern cybersecurity standards and data protection mandates (Wan et al., 2022; Masella & Gramaglia, 2022).

CONCLUSION

This article demonstrates the transformative potential of an automated, machine learning-based approach to threat correlation in cybersecurity. By integrating advanced algorithms such as clustering and classification algorithms like neural networks, the proposed model addresses key limitations of traditional, manual threat detection methods, offering substantial improvements in response times, detection accuracy, and resource optimization. Automated threat

correlation enables organizations to process large amounts of “big data” in real-time, effectively identifying sophisticated threats and minimizing human error. This process reduces the burden on cybersecurity teams and enhances the ability to detect complex, multi-step attacks that manual methods might miss.

The demand for agile and accurate detection methods continues to grow despite the fast evolution in cybersecurity threats, emphasizing the value of this machine learning framework. In the current cybersecurity environment where threats can have wide-reaching economic and institutional impacts, automated threat correlation provides a timely, efficient, and scalable solution. This framework could shape the future of cybersecurity practices, empowering organizations to preemptively defend against attacks and ensuring greater resilience in a digitally dependent world.

FUTURE WORK

For future development of this automated threat correlation framework, several improvements and areas of research could enhance its efficacy and adaptability. One promising avenue involves integrating more advanced machine learning algorithms, such as deep learning models. Deep learning, particularly through models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has shown substantial promise in identifying complex patterns and anomalies within big datasets, making it a viable choice for handling increasingly sophisticated cyber threats. Using these models, the framework could better manage the dynamic nature of threats and improve its ability to detect and correlate multi-step attack patterns.

To complement algorithmic advancements, expanding data sources to include real-time threat intelligence feeds, open-source intelligence (OSINT), and diverse network data from endpoint devices can strengthen the model. Such data sources would provide a broader understanding of potential threats and enable the framework to adapt more quickly to emerging attack vectors. Continued research on real-time threat correlation methods is essential, as current frameworks often struggle with latency and scalability when handling vast, high-velocity datasets.

Further studies are also needed to develop models that can accurately detect and correlate increasingly complex attack scenarios, such as those involving advanced persistent threats (APTs) and multi-stage intrusions.

REFERENCES

[1] Ahsan, Mostofa, Kendall E. Nygard, Rahul Gomes, Md Minhaz Chowdhury, Nafiz Rifat, and Jayden F Connolly. 2022. "Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning—A Review" Journal of Cybersecurity and Privacy 2, no. 3: 527-555. https://doi.org/10.3390/jcp2030027

[2] Aminu, Muritala & Akinsanya, Ayokunle & Oyedokun, Oyewale & Dickson, Apaleokhai & Dako,. (2024). Enhancing Cyber Threat Detection through Real-time Threat Intelligence and Adaptive Defense Mechanisms. International Journal of Computer Applications Technology and Research. 13. 11-27. 10.7753/IJCATR1308.1002

[3] Aslan, Ömer & Aktug, Semih & Ozkan Okay, Merve & Yılmaz, Abdullah & Akin, Erdal. (2023). A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions. Electronics. 12. 1-42. 10.3390/electronics12061333.

[4] Azeem M, Javaid S, Khalil RA, Fahim H, Althobaiti T, Alsharif N, Saeed N. (2023). Neural Networks for the Detection of COVID-19 and Other Diseases: Prospects and Challenges. Bioengineering (Basel). 2023 Jul 18;10(7):850. doi: 10.3390/bioengineering10070850. PMID: 37508877; PMCID: PMC10416184.

[5] Cai, Li & Zhu, Yangyong. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal. 14. 10.5334/dsj-2015-002.

[6] Chao Liu; Zhaojun Gu; Jialiang Wang. (2021). A Hybrid Intrusion Detection System Based on Scalable K-Means+ Random Forest and Deep Learning.

https://ieeexplore.ieee.org/abstract/document/9437227

[7] Cybersainik. (2021). Threat Correlation: A Comprehensive Guide. https://cybersainik.com/threat-correlation-a-comprehensive-guide/#:~:text=Threat%20correlation%20is%20using%20correlation,events%20to%20identify%20potential%20threats.

[8] Dalal, S., Manoharan, P., Lilhore, U.K. et al. (2023). Extremely boosted neural network for more accurate multi-stage Cyber attack prediction in a cloud computing environment. J Cloud Comp 12, 14 (2023). https://doi.org/10.1186/s13677-022-00356-9

[9] Dhongade, Gauri & Chandrakar, Dr & Khande, Rajeshree. (2024). Enhancing Cyber Security: A Study of Data Preprocessing Techniques for Cyber Security Datasets. International Journal of Scientific Research in Science and Technology. 11. 71-75. 10.32628/IJSRST2411427.

[10] Dol Aher, Sunita & J., Lobo. (2012). Combination of Clustering, Classification & Association Rule-based Approach for Course Recommender System in E-learning. International Journal of Computer Applications. 39. 8-15. 10.5120/4830-7087.

[11] Franklim Arévalo, Paolo Barucca, Isela-Elizabeth Téllez-León, William Rodríguez, Gerardo Gage, Raúl Morales. (2022). Identifying clusters of anomalous payments in the Salvadorian payment system. Latin American Journal of Central Banking. https://doi.org/10.1016/j.latcb.2022.100050.

[12] Haas, Steffen & Wilkens, Florian & Fischer, Mathias. (2019). Efficient Attack Correlation and Identification of Attack Scenarios based on Network-Motifs. 10.48550/arXiv.1905.06685.

[13] IBM (2024). IBM Report: Escalating Data Breach Disruption Pushes Costs to New Highs https://newsroom.ibm.com/2024-07-30-ibm-report-escalating-data-breach-disruption-pushes-costs-to-new-highs#:~:text=Hacking%20the%20clock%20with%20AI,those%20disclosed%20by%20an%20attacker.

[14] J. Pavithra, S. Selvakumara Samy. (2022) A Comparative Study on Detection of Malware and Benign on the Internet Using Machine Learning Classifiers. Mathematical Problem Engineering. https://doi.org/10.1155/2022/4893390

[15] Kiran Maharana, Surajit Mondal, Bhushan Kumar Nemade. (2022). A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings. https://doi.org/10.1016/j.gltp.2022.04.020

[16] Kumar, Mukesh. (2012). Evaluating the performance of apriori and predictive apriori algorithms to find new association rules based on the statistical measures of datasets. IJERT "International Journal of Engineering Research and Technology". 1. 1-5.

[17] Li, Jingwen, Jianyi Liu, and Ru Zhang. 2024. "Advanced Persistent Threat Group Correlation Analysis via Attack Behavior Patterns and Rough Sets" Electronics 13, no. 6: 1106. https://doi.org/10.3390/electronics13061106.

[18] Maosa, Herbert, Karim Ouazzane, and Mohamed Chahine Ghanem. 2024. "A Hierarchical Security Event Correlation Model for Real-Time Threat Detection and Response" Network 4, no. 1: 68-90. https://doi.org/10.3390/network4010004

[19] Massella M, Dri DA, Gramaglia D. (2022). Regulatory Considerations on the use of Machine Learning-based tools in Clinical Trials. Health Technol (Berl). 2022;12(6):1085-1096. doi: 10.1007/s12553-022-00708-0. Epub 2022 Nov 7. PMID: 36373014; PMCID: PMC9638313.

[20] Mayra Macas, Chunming Wu, Walter Fuertes. (2024). Adversarial examples: A survey of attacks and defenses in deep learning-enabled cybersecurity systems. Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2023.122223.

[21] Michal Tonhausera & Jozef Ristveja. (2023). TRANSCOM 2023: 15th International Scientific Conference on Sustainable, Modern and Safe Transport Cybersecurity Automation in Countering Cyberattacks. Transportation Research Procedia

[22] Norsyafawati, Fatin & Sabri, Mohd & Md Norwawi, Norita & Seman, Kamaruzzaman.

(2011) Identifying False Alarm Rates for Intrusion Detection System with Data Mining. https://www.researchgate.net/publication/264880778_Identifying_False_Alarm_Rates_for_Intrusion_Detection_System_with_Data_Mining

[23] Ovais Naseem. 2024. Ensuring Data Accuracy in Machine Learning Models. Data-Driven Investors. https://www.datadriveninvestor.com/2024/08/28/ensuring-data-accuracy-in-machine-learning-models/

[24] Potter, Kaledio & Doris, Lucas. (2024). AI-POWERED THREAT DETECTION AND INCIDENT RESPONSE SYSTEMS. Cybersecurity.

[25] Sarker, I.H., Kayes, A.S.M., Badsha, S. et al. (2020). Cybersecurity data science: an overview from a machine learning perspective. J Big Data 7, 41. https://doi.org/10.1186/s40537-020-00318-5

[26] Salem, A.H., Azzam, S.M., Emam, O.E. et al. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. J Big Data 11, 105 (2024). https://doi.org/10.1186/s40537-024-00957-y

[27] Sameera, Nerella & Jyothi, M.Siva & K.Lakshmaji, & Neeli, V S R Pavan Kumar. (2023). Clustering-based Intrusion Detection System for effective Detection of known and Zero-day Attacks. Journal of Advanced Zoology. 44. 969-975. 10.17762/jaz.v44i4.2423.

[28] Sharma, Vinod. (2022). A Study on Data Scaling Methods for Machine Learning. International Journal for Global Academic & Scientific Research. 1. 10.55938/ijgasr.v1i1.4.

[29] Shahid Tufail, Hugo Riggs, Mohd Tariq, and Arif I. Sarwat. (2023). Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms. https://doi.org/10.3390/electronics12081789

[30] Siraj, Fakiha. (2020). CYBERSECURITY SITUATION AWARENESS 1 Effectiveness of Security Incident Event Management (SIEM) system for Cyber Security Situation Awareness. Indian Journal of Forensic Medicine & Toxicology. 14. 10.37506/ijfmt.v14i4.11587.

[31] Statista. (2023). The annual amount of financial damage caused by reported cybercrime in the U.S. 2001-2023. https://www.statista.com/statistics/267132/total-damage-caused-by-by-cybercrime-in-the-us/#:~:text=Annual%20amount%20of%20financial%20damage,cybercrime%20in%20U.S.%202001%2D2023&text=In%202023%2C%20the%20monetary%20damage,of%2012.5%20billion%20U.S.%20dollars.

[32] Stellar Cyber (2021). Stellar Cyber's Open XDR Debuts AI-Powered Incident Correlation to Reveal and Stop Cyberattacks Faster https://stellarcyber.ai/news/press-releases/stellar-cyber-debuts-ai-powered-incident-correlation-to-reveal-and-stop-cyberattacks-faster/

[33] Sumathi P., Srinivasan. K, Meenu Parames. P, Sathya Murthy. D, Hari Krishnan B. (2024). AI-Powered Defense Against Phishing Threats for Enterprises. https://doi.org/10.22214/ijraset.2024.60395

[34] Taye, Mohammad Mustafa. 2023. "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications, and Future Directions" Computers 12, no. 5: 91. https://doi.org/10.3390/computers12050091

[35] Thapliyal, Vikalp & Thapliyal, Pranita. (2024). Machine Learning for Cybersecurity: Threat Detection, Prevention, and Response. Darpan International Research Analysis. 12. 1-7. 10.36676/dira.v12.i1.01.

[36] Trayi Chaganti & Rohith Tadi. (2022). Comparison of Machine Learning Algorithms for Anomaly Detection in Train's Real-Time Ethernet using an Intrusion Detection System. Faculty of Computing, Blekinge Institute of Technology. https://www.diva-portal.org/smash/get/diva2:1707593/FULLTEXf

[37] Wan, Wai & Tsimplis, Michael & Siau, Keng & Yue, Wei & Nah, Fiona & Yu, Gabriel. (2022). Legal and Regulatory Issues on Artificial Intelligence, Machine Learning, Data Science, and Big Data. 10.1007/978-3-031-21707-4_40.

[38] Wasyihun Sema Admass, Yirga Yayeh Munaye, Abebe Abeshu Diro. (2024). Cyber security: State of the art, challenges, and future directions. Cyber Security and Applications. https://doi.org/10.1016/j.csa.2023.100031.