International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 74-80, January-2024

https://doi.org/10.47001/IRJIET/2024.801010

CyMac: Diving Deep into the Application of Machine Learning Algorithms in Cyber Security

[1]*Bishwajit Das, [2]Nikita Yadav, [3]Deepa Chauhan, [4]Sanju Gupta

[1,2,3,4]Department of Computer Science & Engineering, Chhatrapati Shivaji Maharaj University, Navi Mumbai, India

***Abstract* - Machine learning has emerged as a climatic technology in contemporary and prospective cyber threat intel systems, with numerous jurisdictions seamlessly integrating it into their operations. However, the current state of machine learning in cyber defence is still in its early stages, foreshadowing a noticeable unexplored research territory and practical implementation. This paper marks the initial endeavour to offer a comprehensive understanding of machine learning within the entire spectrum of cybersecurity jurisdictions, catering to potential end users with enthusiasm in this field of study. This paper aims to serve as a source of inspiration for significant advancements in ML within the cyber defence zone, laying the groundwork for the broader adoption of ML mitigations to safeguard present and heuristic systems.**

Keywords: Machine Learning, Cybersecurity, Enthusiasm technology, Jurisdictions, Supervised, Un-Supervised, Domain, Elucidating, Cyber Vulnerabilities, Phishing Attacks, Intrusion Detection Systems, SNORT, Host Intrusion Detection Systems.

I. INTRODUCTION

As modern information systems continue to grow in complexity and generate an ever-expanding flow of big data, the advantages of Machine Learning (ML) in the realm of cybersecurity are becoming increasingly apparent and widely acknowledged. More specifically, within the realm of cybersecurity, Machine Learning (ML) methods have already been applied to tackle a broad range of real-world tasks, a trend that has been particularly accelerated with the advent of deep learning. Moreover, in cybersecurity, Machine Learning (ML) is rightfully acknowledged as a technology facilitator, showcasing significant promise in enhancing threat detection and response capabilities and bolstering the security of critical infrastructure. Digital threats are indeed in a perpetual state of evolution, and as per Gartner's prediction, by 2025, attackers may find their capabilities insufficient to directly endanger or cause harm to humans. Consequently, decision-makers must possess a comprehensive grasp of the (i) advantages, (ii) limitations, and (iii) hurdles associated with a cybersecurity solution before giving their approval for practical implementation.

This paper is designed to be accessible to all readers, regardless of their level of technical expertise, within the context of Machine Learning (ML) in cybersecurity.

Figure 1 illustrates timestamp data concerning a particular date. The x-axis denotes the matching popularity, while the y-axis signifies the corresponding popularity within the range of 0 (minimum) to 100 (maximum). We have drawn this graph to show how machine learning has gained growth in cybersecurity for controlling vulnerabilities and in the future how it will take more growth in cybersecurity. Machine Learning in cybersecurity helps the cybersecurity analyst to easily detect any threats in the software or any system so that the analyst can easily take action against it.

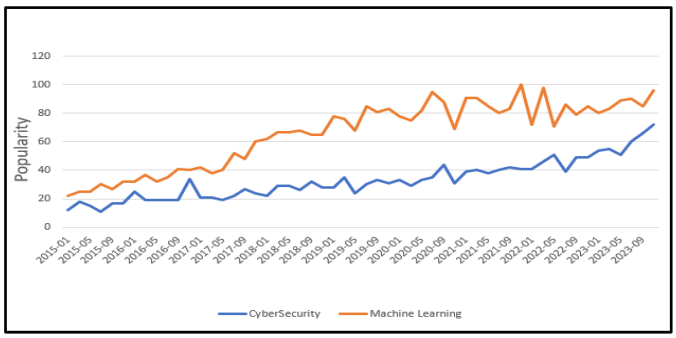

Figure 1: Increasing Popularity of Machine Learning in Cybersecurity

We commence in Section 2-the Mild Initiation to Machine Learning by introducing the fundamental concepts of the ML paradigm without delving into notations. Moving on to Section 3, we delve into the Scope and targets of ML in cybersecurity. Continuing to Section 4, we illuminate different cybersecurity defence strategies.

To establish the foundation for our paper within the context of Machine Learning in cybersecurity, we initially introduce key concepts in a simplified manner, ensuring accessibility to readers of all backgrounds. Our objective is to acquaint readers with established terminology and the typical categories of prevailing ML techniques. Subsequently, we outline the scope of this paper and identify its intended readership. Additionally, we underscore the distinctions

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 74-80, January-2024

https://doi.org/10.47001/IRJIET/2024.801010

between our approach and prior research efforts, specifically within the realm of Machine Learning (ML) in cybersecurity.

II. MILD INITIATION TO MACHINE LEARNING

In the context of Machine Learning (ML) in cybersecurity, the objective is to create automated systems capable of making decisions autonomously. This model encompasses the knowledge acquired during the training phase and is designed to execute a decision-making function when presented with 'future' data. Before deploying an ML model in an operational cybersecurity environment, it is imperative to evaluate its performance thoroughly.

Supervised techniques explicitly demand the availability of labeled training data. In certain instances, these labels may be generated naturally, while in other cases, acquiring labels necessitates dedicated manual verification efforts. In contrast, unsupervised approaches either operate without the need for labels or involve minimal supervision, particularly within the context of Machine Learning (ML) in cybersecurity. For example, Machine Learning (ML) in cybersecurity, the reinforcement learning constructs the ML model using a fully automated feedback mechanism.

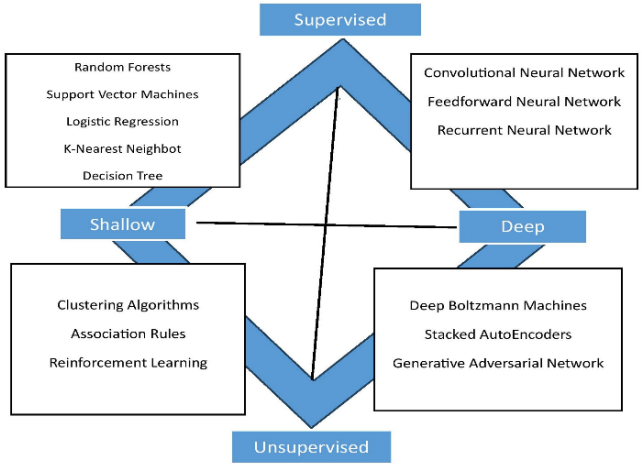

Figure 2: Illustrates common Machine Learning (ML) algorithms

An algorithm is categorized as 'deep' if it is based on neural networks; otherwise, it is considered 'shallow.' Algorithms that necessitate labeled data are employed in 'supervised' tasks, while those applicable without labeled data can also be utilized in 'unsupervised' tasks. These distinctions are relevant within the context of ML in cybersecurity as well.

It's customary to distinguish between 'positives' (representing malicious activities) and 'negatives' (representing benign activities). Evaluating performance involves considering both the accurate predictions (i.e., True Positives and True Negatives) and the erroneous predictions (i.e., False Positives and False Negatives) generated by a specific model, particularly within the context of Machine Learning (ML) in cybersecurity. It's important to highlight that appraisal assessment applies to ML models rather than ML methods themselves. Depending on the particular context, such as the training data, the ML algorithm used, and its various parameters, an ML method may produce multiple ML models, each exhibiting distinct performance characteristics. This applies to the domain of Machine Learning (ML) in cybersecurity as well.

III. SCOPE AND TARGET

The focus of our paper is to establish a connection between the academic research and practical application of Machine Learning (ML) within the field of cybersecurity. Our paper is designed to be obtainable by any reader interested in understanding the intersection of ML and cybersecurity.

Specifically, we cater to the following three categories of readers within the cybersecurity domain:

- Decision-makers, including Corporate Executives and Chief Information Security Officers, who require insights into the current state-of-the-art. This paper aims to enable more informed decisions regarding the adoption of the incorporation of Machine Learning (ML) into current systems to improve the effectiveness of security operations.
- Security professionals, such as security consultants, administrators, and digital forensics experts, need a comprehensive understanding of operational considerations and the practical applications of ML in cybersecurity. Such knowledge is vital for effectively addressing cybersecurity challenges.
- Engineers with an enthusiasm for creating innovative ML solutions for cybersecurity, enhancing existing ML systems, or addressing their limitations. The paper's identification of outstanding concerns and challenges should prove to be a useful roadmap for ML's future progress in the cybersecurity space.

1) Cybersecurity Defence Strategies:

In cybersecurity, defence strategies play a crucial role in safeguarding information, information systems, and networks against cyber attacks or unauthorized intrusions. These strategies are responsible for both proactively preventing data incidents and actively monitoring and responding to threats, which encompass any unlawful activities that pose harm to a network or individual systems.

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 74-80, January-2024

https://doi.org/10.47001/IRJIET/2024.801010

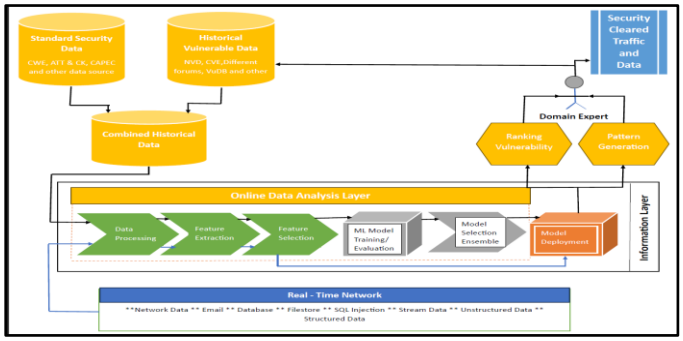

Figure 3: Flowchart of Cybersecurity Defence Strategies

IV. CYBER-SECURITY: UTILIZING MACHINE LEARNING FOR THREAT DETECTION

In the domain of cybersecurity, the security lifecycle encompasses three key processes: prevention, detection, and reaction. Consequently, the majority of security mechanisms, including those leveraging machine learning, primarily concentrate their efforts on threat detection. Two detection approaches complement each other: misuse-based methods are highly accurate but can only identify threats that match known patterns, Approaches based on anomalies typically produce but offer better potential for detecting novel and previously unseen attacks. As a response to this challenge, data-driven solutions, including ML, came into play within the detection systems. These solutions not only reduced the manual effort required but, in some instances, even surpassed the performance of traditional, manually crafted detection methods.

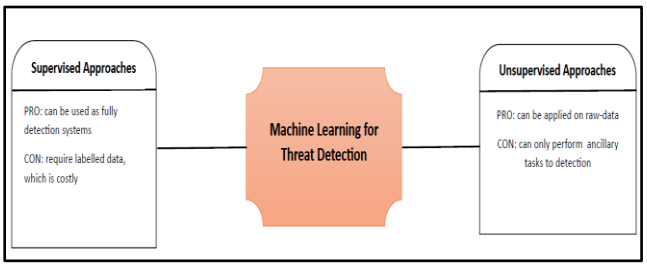

Figure 4: Advantages and Disadvantages of Supervised and Unsupervised ML in Cyber Threat Detection

To maintain focus and relevance, we categorize this section into three major areas of cyber detection: network intrusion detection, malicious software detection, and phishing detection.

1) Utilizing Machine Learning for Detecting Network Intrusions:

One of the primary areas of concern in modern enterprises within the field of cybersecurity is Intrusion Detection, which is facilitated through Intrusion Detection Systems (IDS). IDS can be categorized into two main types: network Intrusion Detection Systems (NIDS) that examine network-level activities, and Host Intrusion Detection Systems (HIDS) that assess activities at the individual host level. Over the past decade, numerous machine learning (ML) solutions have been introduced to enhance the effectiveness of Network Intrusion Detection Systems (NIDS). These solutions have been explored extensively in both scientific literature and patents. Signature-based detection is quite successful in identifying known attacks and relies on specific knowledge about previously observed intrusion patterns. An example of a widely recognized signature-based ID is SNORT.

Unsupervised ML methods are particularly valuable within the framework of cybersecurity because of the challenges associated with acquiring labeled data for entire networks. Among these approaches, we want to emphasize the outcomes achieved using clustering methods. The integration of ML-enabled the detection of 12 malicious hosts, whereas the commercial NIDS only detected 3.Unsupervised methods can play a vital part in aiding the manual creation of regulations for NIDS based on usage.

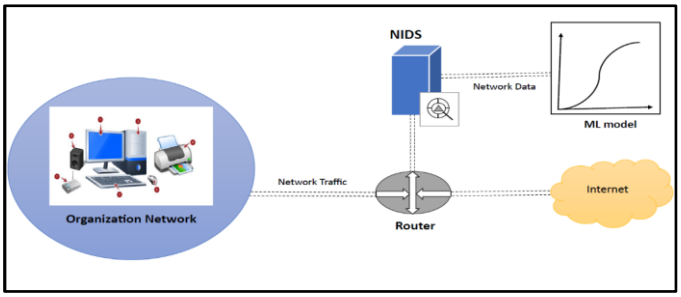

Figure 5: Common Configuration of an ML-NIDS Deployment

Network traffic is routed through a NIDS by the border router, where it undergoes further analysis using ML models.

Conversely, presents contrasting results, with their 'deep' neural network also achieving a 0.96 F1-score, while their shallow decision tree reaches a notably higher 0.99 F1-score.

Table 1

Routing Strategy	Attack payload (# Interest/s)	Range 5-9	Range 10-19	Range 20-49	Range 50
Best Route	%True Positive	95.0%	95.3%	97%	98.3%
	%False Positive	1.0%	1,0%	1.0%	1.0%
Multicast	%True Positive	63.3%			
	%False Positive	1.0%	1.0%	1.0%	1.0%

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 8, Issue 1, pp 74-80, January-2024
https://doi.org/10.47001/IRJIET/2024.801010

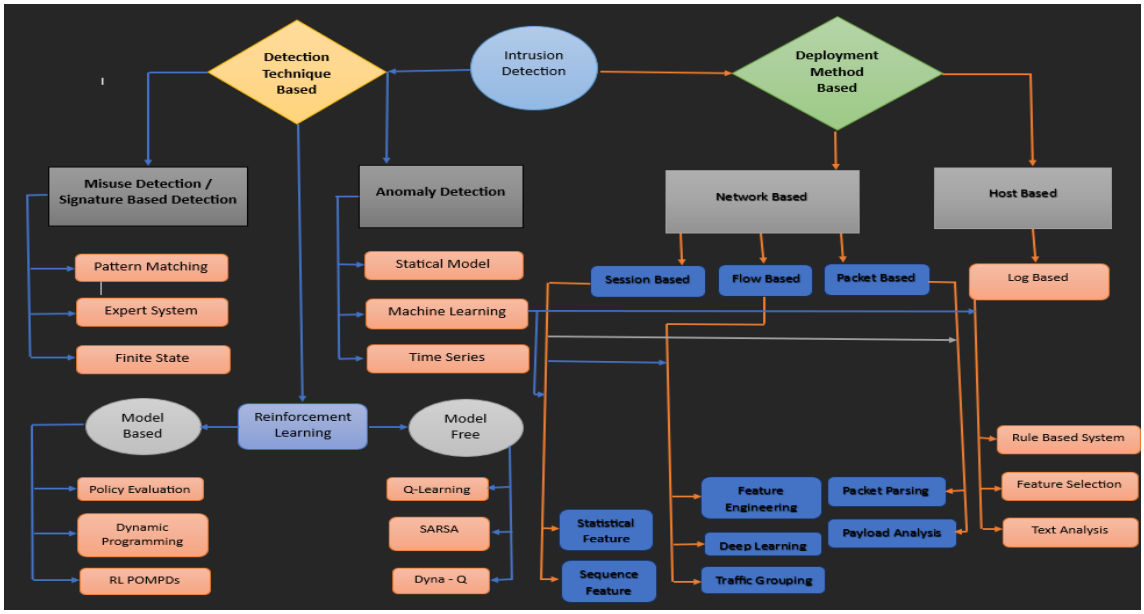

Figure 6: Types of Intrusion Detection Systems

2) Utilizing Machine Learning for the Identifying Malevolent Software in Cybersecurity:

The battle against malicious software, commonly known as malware, represents one of the prominent challenges in the realm of cybersecurity. The malware primarily impacts individual devices, necessitating its detection through the analysis of host-level data, often facilitated by Host Intrusion Detection Systems (HIDS). For over two decades, Windows OS has been the most frequent target for malware due to its widespread usage. Malware detection employs two primary methods of analysis: static and dynamic.

It may be somewhat improved through the application of Machine Learning techniques. However, it's important to note that static malware detection methods are susceptible to evasion.

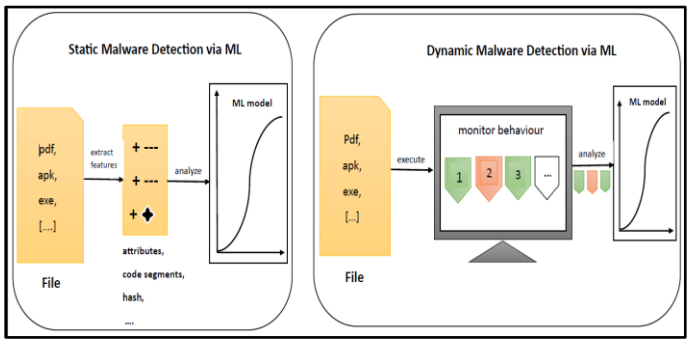

Figure 7: Malware Detection Using ML Approaches

In static analysis, an ML model extracts and examines the properties of a file. In dynamic analysis, the file is executed, and its behavior is continuously monitored, and then assessed by an ML model.

Finally, it's feasible to merge static and dynamic analyses with the aid of machine learning, as demonstrated in EC2, where unsupervised and supervised machine learning is combined in order to find new malware for Android malware, yielding a detection rate exceeding 90%.

3) Machine Learning for Detecting Phishing Attempts in Cybersecurity:

In the modern organizational context, the early identification of attempted phishing holds immense significance, and machine learning (ML) plays an essential role in achieving this goal. The primary differentiation between these two approaches lies in the analyzed data: for phishing websites, the analysis commonly involves the URL, HTML code, or visual representation of the webpage, while for phishing emails; the focus is on scrutinizing the email's text, header, or attachments. To provide a more comprehensive understanding, we offer a detailed description of these applications illustrated in Fig. 7.

V. EXPANDING THE HORIZONS OF MACHINE LEARNING IN CYBERSECURITY: BEYOND DETECTION

Apart from threat detection, machine learning can fulfill numerous supplementary functions within the realm of cybersecurity. In contemporary settings, vast volumes of data are continuously generated, emanating from a wide array of sources, which may include the very machine learning models discussed in Section 3. Leveraging machine learning for the analysis of such data can yield valuable insights, ultimately enhancing the security of digital systems.

© 2024-2017 IRJIET All Rights Reserved

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 74-80, January-2024

https://doi.org/10.47001/IRJIET/2024.801010

1) Alert Handling:

- Alert Screening: In cybersecurity, it's important to recognize that alerts aren't inherently indicative of malicious activity, and a considerable portion of alerts may be false positives.
- Cybersecurity Alert Ranking: Machine learning proves advantageous in this context as it can autonomously "learn" the most pertinent criteria for ranking alerts with minimal human supervision.
- Alert Consolidation: Effectively handling a deluge of alerts in the realm of cybersecurity involves the process of aggregating akin alerts and subsequently exploring correlations within these groups to uncover relevant causal relationships crucial for security-related tasks.

2) Data Examination at the Raw Level:

- Security Operations: The wealth of log data within contemporary information systems underscores the promise of machine learning within the domain of cybersecurity operations.
- Enhanced Labelling Strategies: Numerous threat detection methodologies depend on supervised machine learning, often necessitating extensive volumes of accurately labeled data, a crucial consideration within the domain of cybersecurity.

3) Cybersecurity Risk Evaluation:

- Automated Security Assessment: In the realm of cybersecurity, machine learning proves to be a valuable asset for evaluating vulnerabilities by simulating automated attacks on existing security systems.
- Predicting Compromised Hosts: Within the scope of cybersecurity, machine learning can be harnessed to make predictions regarding the most probable compromised hosts within a specific system.

4) Cybersecurity Intelligence:

- Using Internal Corporate Data: The anticipation of future attack strategies through machine learning can be exclusively accomplished by utilizing internal business data in the cybersecurity domain. Using AI to Enhance Open Source Intelligence (OSINT) in Cybersecurity.

VI. PROSPECTS FOR MACHINE LEARNING IN THE FIELD OF CYBERSECURITY

Moving Forward Positively, this segment underscores the potential transformative advancements on the horizon for machine learning in cybersecurity. While we acknowledge and value every improvement, we firmly assert that bridging the current disparity between research and practical implementation necessitates collaborative efforts from four key stakeholders: regulatory entities, corporate leadership, cybersecurity professionals, and the research community.

VII. REAL-WORLD EXAMPLES & PRACTICAL APPLICATIONS OF MACHINE LEARNING IN CYBERSECURITY

As a last addition to this document, we offer two case studies illustrating concrete and successful applications of artificial intelligence in the realm of cybersecurity.

Many products in the commercial sector assert the incorporation of machine learning into their cybersecurity solutions.

1) Identifying Cache Poisoning Attacks in Named Data Networks:

Situational Context and Hurdles: This case study delves into the renowned ICN methodology of Named Data Networking (NDN). In the NDN approach, a pull-based mechanism is employed, featuring two primary types of packets: Interest (a content request) and Data (the response containing the content). When a user wishes to access specific content, the user (i) specifies the desired content's name (e.g., "/data/video.mp4") within an Interest, (ii) dispatches this Interest through the NDN network, and (iii) subsequently receives the corresponding Data. This Data may originate from the content producer or any intermediary NDN node that stores a copy of said Data.

2) Incorporating Machine Learning and Non-Machine Learning Methods for the Protection of Industry 4.0 Cybersecurity:

Context and Obstacles: This case study underscores the benefits of employing machine learning applications for detecting anomalies in time-series data within the realm of cybersecurity. The rationale behind this approach is that Advanced Persistent Threats (APTs) exploit zero-day vulnerabilities, making them impervious to detection through misuse-based methods, whether driven by humans or data.

This design choice proves especially suitable for practical Industrial Control System (ICS) deployments, offering a triple benefit compared to 'one-size-fits-all' ML architectures. These advantages encompass the following aspects:

1. Individual ML Training models are simpler since they only need to handle a small fraction of the information, leading to improved performance as well as reduced bogus alarms.

© 2024-2017 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 8, Issue 1, pp 74-80, January-2024
https://doi.org/10.47001/IRJIET/2024.801010

2. It permits the combination of diverse algorithms, one for each tailored address a particular issue as well as data type.
3. It enhances the system's adaptability for the future, allowing for individual updates, removals, or replacements of each ML model. This adaptability is crucial within the cybersecurity context.

VIII. CONCLUSION

This paper offers a comprehensive exploration of the role of Machine Learning (ML) in the field of cybersecurity, presenting an executive summary of the advantages, challenges, and forthcoming prospects of ML within this domain. Upon presenting the fundamental ML concepts, the paper briefly outlines their usage in identifying triple categories of major online dangers: network intrusions, phishing, and malware. To tackle these obstacles, cooperation is needed from various domains: governing and authoritative organizations, corporations of leadership, and the broader scientific community, in addition to engineers. In conclusion, we have two examples of case studies exemplifying successful and operational industrial implementations of ML in order to counter cyber threats.

REFERENCES

[1] Nir Kshetri. 2021. Economics of Artificial Intelligence in Cybersecurity. IEEE IT Professional 23, 5(2021), 73–77.

[2] 2021. Darktrace Industrial Uses Machine Learning to Identify Cyber Campaigns Targeting Critical Infrastructure. https://www.darktrace.com/en/press/2017/204/

[3] Last line. 2020. Using AI to Detect and Contain Cyberthreats. Technical Report. https://www.lastline.com/wp-content/uploads/2020/01/Lastline_WP_AI_Done_Right_web.pdf

[4] Mohammad S. Jalali, Michael Siegel, and Stuart Madnick. 2019. Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. Elsevier, The Journal of Strategic Information Systems 28, 1 (2019), 66–82.

[5] Ravi Vijayakumar, Mamoun Alazab, KP Soman, Prabaharan Poornachandran, Ameer Al-Nemrat, and Sitalakshmi Venkatraman. 2019. Deep learning approach for the intelligent intrusion detection system. IEEE Access 7 (2019), 41525–41550.

[6] Camila Pontes, Manuela Souza, João Gondim, Matt Bishop, and Marcelo Marotta. 2021. A new method for OW-based network intrusion detection using the inverse Potts model. IEEE Transactions on Network and Service Management (2021).

[7] Chika Yinka-Banjo and Ogban-AsuquoUgot. 2020. A review of generative adversarial networks and their application in cybersecurity.

[8] Giovanni Apruzzese, Michele Colajanni, Luca Ferretti, Alessandro Guido, and Mirco Marchetti. 2018. On the Electiveness of Machine and Deep Learning for Cybersecurity. In Proc. IEEE International Conference on Cyber Conicts. 371–390.

[9] Daniel S Berman, Anna L Buczak, Jeffrey S Chavis, and Cherita L Corbett. 2019. A survey of deep learning methods for cyber security.

[10] Rakesh M. Verma, Victor Zeng, and HoutanFaridi. 2019. Data quality for security challenges: Case studies of phishing, malware, and intrusion detection datasets. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. 2605–2607.

[11] Yann Le Cun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature 521, 7553 (2015), 436–444.

[12] Kasun Amarasinghe, Kevin Kenney, and Milos Manic. 2018. Toward explainable deep neural network-based anomaly detection. In Proc.

[13] Nisreen Alzahrani and Daniyal Alghazzawi. 2019. A review on Android ransomware detection using deep learning techniques. In Proc. ACM Int. Conf. Manag. Digit. Eco Syst. 330–335.

[14] Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. 2012. Machine learning strategies for time series forecasting. In European business intelligence summer school. 62–77.

[15] Petar Radanliev, David De Roure, Rob Walton, Max Van Kleek, Rafael Mantilla Montalvo, Omar Santos, Peter Burnap, Eirini Anthi, et al. 2020. Artificial intelligence and machine learning in dynamic cyber risk analytics at the edge. SN Applied Sciences 2, 11 (2020), 1–8.

[16] Emilie Bout, Valeria Loscri, and Antoine Gallais. 2021. How Machine Learning changes the nature of cyberattacks on IoT networks: A survey. IEEE Commun. Surv. Tut. (2021).

[17] Joseph Gardiner and Shishir Nagaraja. 2016. On the security of machine learning in malware c&c detection: A survey. ACM Computing Surveys (CSUR) 49, 3 (2016), 59.

[18] Daniele Ucci, Leonardo Aniello, and Roberto Baldoni. 2019. Survey of machine learning techniques for malware analysis. Computers & Security 81 (2019), 123–147.

© 2024-2017 IRJIET All Rights Reserved

ISSN (online): 2581-3048
Volume 8, Issue 1, pp 74-80, January-2024
https://doi.org/10.47001/IRJIET/2024.801010

[19] Tushaar Gangavarapu, CD Jaidhar, and Bhabesh Chanduka. 2020. Applicability of machine learning in spam and phishing email filtering: review and approaches. Artificial Intelligence Review (2020), 1–63.

[20] Asif Karim, Sami Azam, Bharanidharan Shanmugam, Krishnan Kannoorpatti, and Mamoun Alazab. 2019. A comprehensive survey for intelligent spam email detection. IEEE Access 7 (2019), 168261–168295.

Citation of this Article:

Bishwajit Das, Nikita Yadav, Deepa Chauhan, Sanju Gupta, "CyMac: Diving Deep into the Application of Machine Learning Algorithms in Cyber Security" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 1, pp 74-80, January 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.801010

****

© 2024-2017 IRJIET All Rights Reserved