https://doi.org/10.46335/IJIES.2023.8.5.15 **e-ISSN: 2456-3463**
Vol. 8, No. 5, 2023, PP. 68-74

International Journal of Innovations in Engineering and Science, www.ijies.net

Comprehensive Review on Enhancing Role of Machine Learning For Intelligent Data Analysis and Automation in Cyber Security

Yogita H. Dhande[1], Pallavi P. Surwade[2]

[1,2] Assistant Professor
[1,2] G H Raisoni Institute of Engineering and Business Management Jalgon, India,, 425002.

yogita.dhande@raisoni.net

***Received on**: 15 April, 2023* ***Revised on**: 15 May, 2023* ***Published on**: 17 May, 2023*

***Abstract** – Due to the rapid growth of different types of cyber-attacks and threats, traditional security solutions are not sufficient to meet today's security challenges. The use of machine learning techniques is essential to provide automated security systems that are dynamically improved and updated through the analysis of security data. In this article, we give an overview of machine learning technique and its applications, highlighting how they can extract valuable insights from network data and using them for intelligent data analysis and security automation of the network. The future of machine learning in cyber security, according to our research, as well as related research directions, are finally highlighted. Our aim is to analyze current and future applicability of machine learning and related methods in cyber security.*

***Keywords** - Data Analysis, Machine learning, Automation in Cyber security, cyber-attacks, Internet of things (IoT).*

I - INTRODUCTION

Machine learning is finally becoming a fundamental technology for cyber security. Machine learning pre-emotively eliminates cyber threats and strengthens the protective infrastructure through pattern detection, real-time cybercrime mapping, and comprehensive penetration testing. Machine learning algorithms are used on previous datasets and analyses to make better assumptions about system behavior. The system can then regulate its movement and even perform functions not explicitly programmed; a boon for cyber security machine learning that can help with the most common tasks including regression, prediction and classification. ML is the only solution to the data deluge and cyber security talent shortage.

With the passage of time and the significant growth of data technology, many of industries are relying on community connections for good business deals and security matters. Networks and Communications are clearly at risk due to the increase in hacking attacks. Classified human, government and armed networks are more vulnerable to problems, so it is imperative to introduce security measures to prevent large-scale information from being illegally altered, damaged or disclosed. Intrusion detection is a key component of network security with the ability to investigate community activities and detect intrusions/attacks. The analysis was performed using Weka [1]. The dataset provided in Weka produced 10 different models in a 10-fold cross-validation. A weighted average is then calculated and displayed as the final result.

For the evaluation, a limited number of standard classifiers are taken into account. [2]

There are several different methods.

https://doi.org/10.46335/IJIES.2023.8.5.15 e-ISSN: 2456-3463
Vol. 8, No. 5, 2023, PP. 68-74

- Naive Bayes - a probabilistic classifier based on Bayes' theorem.
- Random Forest-A Ml is based on a decision tree algorithm.
- OneR - Evaluates rules one by one and selects the best or optimal set of rules.
- Basic implementation of J48-C4.
- Decision tree algorithm.

I. Spam Detection and Phishing and Filtering

Traditional anti-phishing and anti-spam methods such as word filters, IP blacklists, message filtering, and sender reputation mapping are not effective in combating phishing and spam. To automatically and effectively detect and classify phishing emails, advanced solutions leverage hundreds of input features, analyze syntax using NLP, and use visual analytics. Additionally, these methods can find content that organizations want to block.

Across all industries, the following are currently being considered and used:

- Real-time anti-phishing systems based on classification algorithms and natural language processing capabilities.
- Stack model integrating XG Boost, Gradient Boosted Selection Trees and Light GBM, using HTML and URL attributes of phishing web pages.
- A phishing detection method that uses ML algorithms and features to differentiate phishing sites from real ones.
- Detection of phishing sites using reinforcement learning algorithms and neural networks. Uses reinforcement learning algorithms and neural networks to detect phishing websites. These methods use Monte Carlo algorithms and risk principles.
- A spam detection and identification system using random weighting networks and GA.
- Social media analysis and bionic computing to identify spam profiles on Twitter. Spammer detection using modified k-means [3] using the Levy-flight firefly algorithm and chaotic maps.

II. **Denial of Service (DOS) or Distributed Denial of Service (DDOS)**

The goal of a Denial of Service (DoS) or Distributed Denial of Service (DDoS) attack is to flood a server, service, or network with traffic Internet in order to disrupt its normal functioning.

In a DDoS attack, multiple machines are used to send malicious traffic to its target. These machines may be part of a botnet, which is a collection of infected computers that are infected with malware and are under the control of a single attacker.

DDoS attacks are exceptionally famous and destructive in modern networks for two individual reasons. First, modern security appliances have evolved to prevent some common DoS attacks. Second, the DDoS attack appliance has become the affordable and easy to deploy [4]. For ML based DDoS detection, it is common to use unsupervised learning on untagged data and group traffic based on similarity and dissimilarity of traffic characteristics.

The clustering technique, also known as automatic clustering, allows us to find crucial outliers that deviate from other data analysis points; anomalous operation Some of these machine learning algorithms include K-means clustering, Local Outlier Factor and isolation forest. As cyber threats continue to grow and develop rapidly, security providers are doing everything they can to support defenses. Relying on human intellectuality is not the best solution when the traffic is heavy, the route or behavior in the structure is different.

Machine learning is the great solution to the emergence of zero-day attacks and vulnerabilities, as attacks typically focus on certain affected areas of a device or often have similar intent. Affected areas and folders in computing devices can be inspected by the ML algorithm as it compares the impact of devices or files to older instances of vulnerabilities and zero-day attacks before classifying them as legitimate or malicious. There are many machine learning classification algorithms that can give us the desired results.

.

- Algorithms like decision tree classification and random forest classifiers are very good at interpreting data as legitimate or malicious, but require a lot of input data. For anything more important, it's a small cost, since the data needed for the algorithm is probably readily available on the device they're looking at.
- A hybrid of SVM models and ConvNets with Outlier-based intrusion detection systems can predict malicious zero-day attacks with high accuracy (Using Deep Learning techniques for Effective Zero Day attack detection)
- Classification methods such as K nearest neighbours offer good efficiency in detecting zero-day attacks and predicting vulnerabilities.

https://doi.org/10.46335/IJIES.2023.8.5.15 e-ISSN: 2456-3463
Vol. 8, No. 5, 2023, PP. 68-74

- It comes with extensive data pre-processing required by the algorithm to clearly define the legitimacy of a file (techniques for examining malware and fighting zero-day attacks)
- Clustering methods (such as than K-means clustering) are considered more efficient than basic clustering More useful clustering methods do not give legitimacy to a document if K denotes clustering. It predicts legality based entirely on how the algorithm processes each file. (Review Malware and Zero Day Attack Techniques) When accuracy is 80% or better, one can safely accept the use of ML in
- Detection of Zero Day vulnerabilities is extremely important. When implemented successfully, the algorithm is actually efficient enough to delete malicious files as soon as they are launched.

III. Fundamental Approaches to Malware Detection

Every network security product should include effective, robust, and scalable malware detection capabilities. In the early days of the Internet, cyber threats were relatively rare, and Based on the data they collect about the object, malware detection modules determine if the object is a threat. We may collect this data at different stages: - Pre-execution stage data includes everything we know before the file is executed. Increasingly sophisticated malware and the rapid growth of the Internet have made manual detection rules impractical, necessitating new and more advanced protection techniques. As part of malware detection and classification, antimalware companies leverage machine learning, a field of computing that uses algorithms to perform operations such as image recognition, searching, and taking of decision. There are several types of data that can be used to detect malware today, including host-based, network-based, and cloud-based solutions. [5]

Other roles of ML in Cyber security:

Today's environment constantly generates large volumes of data that can derive from any sources, along with the ML models. Analyzing this (abundant) machine learning data can provide insights that further enhance digital security. Some machine learning functions are described as below:

Alert Management

It is well known (with or without ML) that it is impossible to develop a "perfect" diagnosis. Therefore, the results of detection are often in the form of a warning, to prevent from making a decision based on incorrect guesses. Based on these alerts (for example, relevance, relevant owners, or numbers) may act accordingly.

However, today's environment produces thousands of notifications per hour [6, 11], which makes it difficult to categorize notifications. To address this, the ML can be used to filter, prioritize and even consolidate alerts across multiple scenarios.

Alert filtering

Alerts are not necessarily offensive in meaning, and most alerts correspond to false alerts. As it is inefficient and annoying to receive notifications of so many irrelevant notifications, the ML can help filter out suspicious notifications, example Using a filter to specifically report false alarms generated by ML-NIDS [10]: Stopping botnet traffic really reduces warning time by 75% with poor performance.

Raw Data Analysis

The cyber security field must deal with heterogeneous systems, each producing many different types of raw data (such as logs, reports, alerts). Such sites represent fruitful ground for machine learning that can use its capabilities to make the most of raw data. In this context, we can distinguish two applications of machine learning: data analysis to support decision making and optimizing the tag for effort, and the use of (anonym zed) data to support the deployment of machine learning.

Because of the richness of data sets in modern databases, ML is promising in terms of security operations. The importance of data validation has been highlighted after several major security incidents involving leaks of confidential information. [12] Beehive [13] is one of the first (unsupervised) machine learning systems focused on extracting information from renamed (custom) files, dynamic host control protocol (DHCP) servers, or virtual private networks (VPNs).

The aim is to provide with all these logs in a non-optimal way: data points that don't affect the "standard" log represent "conditions" where there should be human intervention. Beehive analyzed two weeks of EMC Corporation data logs and identified approximately 800 events, 65% of which were real security threats (malicious criminal or illegal activity). In contrast, the non-ML method outperformed as they found only 8 cases (a return of only 1%). Even without machine learning maintenance, the Beehive requires specific engineering guidance: the most important data from each source must be determined by experienced experts. With

https://doi.org/10.46335/IJIES.2023.8.5.15 e-ISSN: 2456-3463
Vol. 8, No. 5, 2023, PP. 68-74

the advent of deep learning, this problem has been overcome.

A notable example is Deep Log [14], which uses a similar target as Beehive to analyze different databases (like Hadoop or OpenStack logs). Deep Log achieved great results in the test environment, only 1% of the available data with a detection rate of around 100% after training.

Security Analysis

While preventing all cyber-attacks is an unattainable goal, systems can be strengthened by focusing on their vulnerabilities, parameter estimation.

ML can be an effective tool for vulnerability assessment by automatically "hacking" existing security systems.

For example, Conventional methods using additive learning for synthetic attacks found the false positives of same number in half the manual analysis., achieving 90% faster than the random stop technique. Recently, Ref. [6] uses deep learning techniques to automatically evade and amplify machine learning-based botnet detectors. Similarly, evaluate the vulnerability of data for ML-crafted SQL injection attacks using [20]. In fact, it has been proposed to use a special machine learning supported platform to perform all these measurements.

According to a recent study [21], the penetration testing potential of machine learning has yet to be exploited.

Estimation of compatibility indicators

ML can be used to envision the most affected hosts on a given system. The authors of the application studied the business environment using machine learning to analyze data from different sources, such as behavior of each host and the complete network, such as endpoint protection tools (McAfee) and even broadcast private information of a host [22] on certain users. The results of the survey show that visiting "business" websites is the primary indicator of hosting attempts (about 30%), with "travel" in second place (about 15%) - which is interesting.

During operation one possibility is to combine machine learning with honey pots (for a different version than in, Ref. [23]): Ref. [24] used this technique to determine which host might have been infected by the malicious botnet.

Finally, Facebook uses machine learning to identify fake accounts by associating variables [25], allowing us to reduce this stress by almost 30%.

Threat Intelligence

Here, the main function is to collect and analyze data to predict new attacks. This is a powerful tool for proactive security protection. But we see that the defense industry importance of the is hypothetically around the value of the product:

The ML approach to cyber threat intelligence must therefore be adjusted to protect the most important data However, machine learning application threats can use internal or else external data (or both) Internal services. Future attack strategies prediction by ML can be reliable on the dataset. For example, [27] uses machine learning to generate alerts that correspond to previous cyber attacks, and then uses these alerts to learn attack behavior, perhaps using other machine learning solutions.

For example, SAGE [26] uses machine learning to compress more than 300,000 individual reports from maximum 100 "against graphs" representing steps of all attacks. Another possibility is to use deep learning to "smash" some performance; this allows us to find some bad patterns that will be repeated in future malware: for example, EKLAVIA [30] In fact, this study was 80% successful. Finally, internal as well as external data can be combined: The authors of [31] use malware history data which provided by Symantec to predict future malwares impacts on the company, and the ML solution provides up to 4x more provides more. Estimated on a non-ML basis.

Foreign aid

Machine learning can be used for so-called open source intelligence.

Information provided from security sources such as the CVS or Common Vulnerability Score stored in well-known databases may also be used. For example, author Ref. [30] used machine learning to predict CVS about 1 week before traditional cyber security feeds.

CVS prediction using machine learning can also be done with dark web data, as shown in Ref. [29].

Analyzing the results using a third-party signature (such as Symantec), the proposed machine learning predicted usage of bad data about 40% and about 10% for food. Ref. [28] allows us to determine the value of the attack.

Future and Research Guidelines

Machine learning has become a buzzword in the cyber security industry. As cyber attacks become sophisticated, widespread and targeted, automation is becoming an crucial tool for security professionals. Security teams urgently need more technology to detect user risk and

https://doi.org/10.46335/IJIES.2023.8.5.15 e-ISSN: 2456-3463
Vol. 8, No. 5, 2023, PP. 68-74

malicious behavior, and machine learning holds promise for the future.

Internet security is considered a "zero tolerance zone"; this means that successful attack will lead to a security failure. To avoid detection, cyber attackers are further complicated by the fact that many malware tools are constantly adding new ways to circumvent anti-virus and other threats.

On the other hand, cyber security is in crisis and future research should focus on cyber threat intelligence strategies that can predict significant events and benefits, rather than relying on prevention and mitigation.

There is a worldwide need for systems based on predictive analysis of cyber risks. Machine learning [35] provides 24/7 monitoring and can manage more data than humans. Therefore, necessary machine learning-based functions such as "prediction", "prevention", "hit detection" and "case response" will be beneficial for the success of the automated cyber security system that enables them to achieve the desired results and potential.

Area of the research report

These are:

Presentation: Prevent or stop an attack without damaging it. The primary purpose of protecting tools are protecting business infrastructure from external attacks is. The use of this measure is to protect network files that are changed or modified correctly frequently.

Prediction: Predict the most likely attack, target and method. Predictive analytics is the most effective way an organization or individual can predict risks, threats, vulnerabilities, or other related cyber issues before they affect or ostensibly affect the body.

Detection: Attacks must be identified in order to respond quickly and accurately. This is usually the process of examining the complete security ecosystem to existence of any malicious behavior or vulnerabilities that could affect to be physically damaged. If a threat is properly recognized and mitigated before exploiting an existing vulnerability, before that the prevention should be done .

Response: Solve problems in time to lessees damage and bring back in normal. The Incident response is an effective way to contain and manage the consequences of a crime. Therefore, the term "incident response" is often used to describe how an organization responded and efforts to a data breach or cyber attack event What to do after the violation attack of ("incident").

Together, we can define "cyber security solutions" that prevent cyber attacks, respond automatically to attacks, and respond to malicious or unusual activity through communication. This is why ML can be used as an important technology as it allows cyber security systems to analyze situations and take guidance from them to help prevent similar attacks like changing behavior.

As a result, Machine learning has the potential to improve cyber security by making it smarter, more efficient, more cost-effective and more efficient. Many machine learning methods are generally classified as supervised or unsupervised learning.. The Supervised learning needs information gathering for detecting malicious activities. Unsupervised learning is better for detecting malicious activity that wasn't seen before the attack, because doesn't need information gathering. Therefore, for choosing a learning algorithm suitable for the application depends on data and its quality, will be tough. One of the most difficult problems is to collect data from edge, network, and cloud, normalize it and use it successfully for machine learning.

If the data is not suitable for learning, such as lack of representation, bad features, poor quality, or insufficient for training,, then the machine learning model will not be any more useful or will return invalid or inappropriate results.

Overall, machine learning has become an essential tool for cyber security. Today, it is nearly impossible to implement effective cyber security solutions without the dependence on the machine learning. However, it is difficult to successfully deploy machine learning without understanding, in-depth, and fully understanding the underlying data of Machine learning-based Cyber security systems can identify patterns and learn patterns to help prevent change behavior and repeated cyber attacks.

Also enables network security teams to be more defensive in preventing the threat and respond to data breaches or hacking attacks in real time. So this machine learning based solution can help individuals as well as organizations better allocate their resources to and lessen the time they used working every day.

Therefore, we need to give more attention to the development of data models or good machine learning algorithms that provide good knowledge or understanding of security, and methods that prepare the data that takes into account the raw data of the world to achieve the specific problems that needs Network security.

https://doi.org/10.46335/IJIES.2023.8.5.15 **e-ISSN: 2456-3463**
Vol. 8, No. 5, 2023, PP. 68-74

International Journal of Innovations in Engineering and Science, www.ijies.net

CONCLUSION

In this article, we give a general introduction to machine learning methods for intelligent data analytics and cyber security operations. We briefly describe the capabilities of various machine learning methods to solve problems in various network applications. The success of the machine learning model depends on the data and how well the learning algorithm performs. Finally, we discuss the challenges in this area and directions for future research.

Altogether, we believe our work on machine ML based models, security solutions are useful and guides further research and implementation for researchers and professionals in the cyber security field.

REFERENCES

[1] *Morris, T. H., Thornton, Z., &Turnipseed, I. (n.d.). Industrial Control System Simulation and Data Logging for Intrusion Detection System Research.*

[2] *WP,ML,DDoS_GN2019https://www.cloudflare.com/learning/ddos/ddos-attack-ools/how-to- ddos/*

[3] *M. Hall, E. Frank, J. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, The WEKA data mining software: an update,‖ ACM SIGKDD Explorations Newsletter, 11 (1), 2009, pp. 10–18 (https://www.genienetworks.com/wpcontent/uploads/ 2019/06/WP_ML_DDoS_GN2019.pdf)*

[4] *Kaspersky-Lab-Whitepaper-Machine-Learning ElectronicGiovanni Apruzzese, MircoMarchetti, Michele Colajanni, 2017. Identifying malicious hosts involved in periodic communications. In the IEEE International Symposium on Network Computing Applications. 1–8.*

[5] *AhmetOkutan and Shanchieh Jay Yang. 2019. ASSERT: Attack synthesis and separation with entropy redistribution towards predictive cyber defense. Cybersecurity 2, 1 (2019), 1–18.*

[6] *AreejAlhogail and AfrahAlsabih. 2021. Applying machine learning and natural language processing to detect phishing email. Comput. Secur. 110 (2021), 102414.*

[7] *KristijanVidović, Ivan Tomičić, KarloSlovenec, MiljenkoMikuc, and IvonaBrajdić. 2021. Ranking network devices for alarm prioritisation: Intrusion detection case study. In Proceedings of the IEEE International Conference on Software, Telecommunications and Computer Networks (SoftCOM'21).*

[8] *Yuan-Hsiang Su, Michael Cheng Yi Cho, and Hsiu-Chuan Huang. 2019. False alert buster: An adaptive approach for NIDS false alert filtering. In Proceedings of the ACM International Conference on Big Data. 58–62.*

[9] *Carter Yagemann, Matthew Pruett, Simon P. Chung, KennonBittick, Brendan Saltaformaggio, and Wenke Lee. 2021. {ARCUS}: Symbolic root cause analysis of exploits in production systems. In Proceedings of the USENIX Security Symposium.*

[10] *Steven McElwee, Jeffrey Heaton, James Fraley, and James Cannady. 2017. Deep learning for prioritizing and responding to intrusion detection alerts. In Proceedings of the IEEE Military Communications Conference. 1–5.*

[11] *Ting-Fang Yen, AlinaOprea, KaanOnarlioglu, Todd Leetham,William Robertson, Ari Juels, and EnginKirda. 2013. Beehive: Large-scalelog analysis for detecting suspicious activity in enterprise networks. In Proceedings of the 29th Annual Computer Security Applications Conference. ACM, 199–208.*

[12] *Min Du, Feifei Li, GuinengZheng, and VivekSrikumar. 2017. Deeplog: Anomaly detection and diagnosis from system logs through deep learning. In Proceedings of the ACM SIGSAC Conference on Computer and Comm Security. 1285–1298.*

[13] *Yong Zhang, JieNiu, Guojian He, Lin Zhu, and Da Guo. 2021. Network intrusion detection based on active semi-supervised learning. In Proceedings of the IEEE International Conference on Dependable Systems and Networks. 129–135.*

[14] *Giovanni Apruzzese, AliyaTastemirova, and PavelLaskov. 2022. SoK: The impact of unlabelled data in cyberthreat detection. In the IEEE European Symposium on Security Privacy.*

[15] *G. Apruzzese, M. Andreolini, M. Marchetti, A. Venturi, and M. Colajanni. 2020. Deep reinforcement adversarial learning against botnet evasion attacks. IEEE Trans. Netw. Serv. Manage. (2020).*

[16] *Xiaohan Zhang, Yuan Zhang, Ming Zhong, Daizong Ding, Yinzhi Cao. 2020. Enhancing -the-art classifiers with API semantics to detect evolved android malware. In the ACM SIGSAC Conference on Computer and Communications Security. 757–770.*

[17] *FeargusPendlebury, Fabio Pierazzi, Roberto Jordaney, Johannes Kinder, and Lorenzo Cavallaro. 2019. {TESSERACT}: Eliminating experimental bias in malware classification across space and time. In the 28th USENIX Security Symposium (USENIX Security'19). 729–746.*

[18] *Solomon OgbomonUwagbole,William J. Buchanan, and Lu Fan. 2017. Applied machine learning predictive analytics to SQL injection attack detection and prevention. In the IFIP/IEEE Symposium on Integrated Network and Service Management (IM'17). 1087–1090.*

[19] *Dean Richard McKinnel, TooskaDargahi, Ali Dehghantanha, and Kim-Kwang Raymond Choo. 2019. A systematic literature review and meta-analysis on artificial intelligence in penetration testing and vulnerability assessment. Comput. Electr. Eng. 75 (2019), 175–188*

[20] *Ting-Fang Yen, Victor Heorhiadi, AlinaOprea, Michael K. Reiter, and Ari Juels. 2014. An epidemiological study of malware encounters in a large enterprise. In Proceedings of the ACM SIGSAC Conference on Computer and Communications Security. 1117–1130.*

[21] *Antonio Nappa, ZhaoyanXu, M. ZubairRafique, Juan Caballero, and GuofeiGu. 2014. Cyberprobe: Towards internet-scale active detection of malicious servers. In Proceedings of the Network and Distributed System Security Symposium (NDSS'14). 1–15.*

[22] *Jose Tomas Martinez Garre, Manuel Gil Perez, and Antonio Ruiz-Martinez. 2021. A novel machine learning-based approach for the detection of SSH botnet infection. Fut. Gener. Comput. Syst. 115 (2021), 387–396.*

[23] *TengXu, Gerard Goossen, HuseyinKeremCevahir, Sara Khodeir, Yingyezhe Jin, Frank Li, Shawn Shan, Sagar Patel, David Freeman,and Paul Pearce. 2021. Deep entity classification: Abusive account detection for online social networks. In Proceedings of the USENIX Security Symposium.*

[24] *AzqaNadeem, SiccoVerwer, Stephen Moskal, and Shanchieh Jay Yang. 2021. Alert-driven attack graph generation using S-PDFA. IEEETrans. Depend. Sec. Comput. (2021).*

[25] *Christopher Sweet, Stephen Moskal, and Shanchieh Jay Yang. 2020. On the variety and veracity of cyber intrusion alerts synthesized by generative adversarial networks. ACM Trans. Manage. Inf. Syst. 11, 4 (2020), 1–21.*

[26] *Rebecca S. Portnoff, SadiaAfroz, Greg Durrett, Jonathan K. Kummerfeld, Taylor Berg-Kirkpatrick, Damon McCoy, Kirill Levchenko, and Vern Paxson. 2017. Tools for automated analysis of cybercriminal markets. In Proceedings of the 26th International Conference on World Wide Web. 657–666.*

[27] *Mohammed Almukaynizi, Eric Nunes, Krishna Dharaiya, Manoj Senguttuvan, Jana Shakarian, and Paulo Shakarian. 2017. Proactive identification of exploits in the wild through vulnerability mentions online. In Proceedings of the IEEE International Conference on Cyber Conflict US (CyCon US'17). Institute of Electrical and Electronics Engineers Inc., 82–88.*

[28] *Daniel Arp, Michael Spreitzenbarth, MalteHubner, Hugo Gascon, KonradRieck, and CERT Siemens. 2014. Drebin: Effective and explainable detection of android malware in your pocket. In Proceedings of the Network and Distributed System Security Symposium (NDSS'14), Vol. 14. 23–26.*

[29] *R. Vinayakumar, MamounAlazab, AlirezaJolfaei, K. P. Soman, and PrabaharanPoornachandran. 2019. Ransomware triage using deep learning: Twitter as a case study. In Proceedings of the IEEE Cybersecurity&Cyberforensics Conference. 67–73.*

[30] *Chanhyun Kang, Noseong Park, B. AdityaPrakash, Edoardo Serra, and V. S. Subrahmanian. 2016. Ensemble models for data-driven prediction of malware infections. In Proceedings of the 9th ACM International Conference on Web Search and Data Mining. 583–592.*

[31] *Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. SNComputSci 2(3):1–21.*