Review Article
ISSN (Online): 3048-894X

Journal of Digital Security and Forensics
July-December 2024 1(1), 26–29

A REVIEW OF THE APPLICATIONS OF MACHINE LEARNING IN CYBERSECURITY AND ITS CHALLENGES

Aishwarya Ulhas Desai [1]

[1] Independent Researcher, India

Received 08 May 2024
Accepted 20 July 2024
Published 02 August 2024

CorrespondingAuthor

Aishwarya Ulhas Desai,
aishwaryaudesai@gmail.com

DOI
10.29121/DigiSecForensics.v1.i1.2024.17

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Our study aims to identify use of Machine Learning (ML) to solve cybersecurity problems and its challenges. ML algorithms are used for malicious network traffic detection, phishing Universal Resource Locator (URLs) detection, malware analysis, and more. Based on the review of selected articles, it is inferred that high-quality data sets, efficient feature extraction, proper training, and testing of ML mode can deliver accurate detection model. Even though the integration of ML in cybersecurity can improve threat detection capabilities, there are some challenges, including limited availability of real-world cyber datasets to train the ML models, the vulnerability of the ML model to data poisoning, imbalanced data sets causing biased detection, and data protection laws making hard to collect necessary data to train and test the ML model efficiently.

Keywords: Machine Learning, Cybersecurity, Deep Learning, Cyber-Attack, Malware Detection, Phishing URL, Network Intrusion Detection

1. INTRODUCTION

Machine Learning is an upcoming technology utilized in various domains to improve the efficiency and accuracy of a process. It can solve complex Cybersecurity problems. Cyber threats are evolving and getting more sophisticated day by day. ML technology offers a modern and practical solution to tackle such evolving threats in the Cybersecurity domain. ML methods detect anomalies, recognize patterns, and adapt in real-time to provide accurate detection capabilities. In the following discussion, we will dive deeper into the applications and challenges of ML in the Cybersecurity domain.

How to cite this article (APA): Desai, A. U. (2024). A Review of the Applications of Machine Learning in Cybersecurity and its Challenges. *Journal of Digital Security and Forensics, 1*(1), 26–29. doi: 10.29121/DigiSecForensics.v1.i1.2024.17

2. METHODS

The author searched Google Scholar on 6th June 2024 with the keywords machine learning and Cybersecurity. Six articles focusing on six applications of machine learning were selected and included in our study.

3. DISCUSSION

According to Xin et al. (2018). Machine learning and deep learning methods for cybersecurity, ML can be used in network-based intrusion detection applications. It mentions that an ML model creation contains steps like feature engineering, choosing an appropriate algorithm, training, and testing the ML model for performance. Various Cybersecurity datasets are available for training and testing ML algorithms. Depending on the application, ML algorithms like Support Vector Machine (SVM), K-NearestNeighbor, decision trees, deep belief networks, and neural networks have different accuracy levels. The authors Rao & Swathi (2017). Fast kNN classifiers for network intrusion detection system, identified that the K-NearestNeighbor algorithm's detection accuracy can reach 99.6% and provide high accuracy in Network Intrusion Detection Systems (NIDS) applications.

Modern Threat Actors (TA) use Artificial Intelligence (AI) to create a phishing email or a phishing website that looks exactly similar to the legitimate one. TA can host a credential-harvesting website, and an unsuspecting victim can access and enter the credentials on the phishing site. TA then uses the harvested credentials to conduct further attacks. Falling for a phishing attack can not only impact the user's digital assets but also cause high-value business email and network compromise. Similarly, the TA can host malware on a spoofed website. Downloading and execution of the payload from the spoofed website can compromise the victim's system. Sahingoz, Ozgur Koray, et al. Machine learning-based phishing detection from URLs designed a real-time ML model to detect phishing URLs. The authors used a combination of seven classification and Natural Language Processing (NLP) algorithms. Evaluating a URL for maliciousness and filtering or blocking connections to the URL can protect unsuspecting users from getting phished. The study determined that a high-performing ML model required a high-quality list of feature extractions. The author leveraged NLP-based word vectors and hybrid feature extraction methods. The author evaluated the accuracy rate using seven algorithms: Decision Tree, AdaBoost, K-star, kNN, Random Forest, SMO, Naïve Bayes algorithm, and NLP. The SMO algorithm had the highest accuracy rate of 97.98%, followed by the Decision Tree algorithm at 97.02%.

Ucci et al. (2019). Survey of machine learning techniques for malware analysis, reviewed the application of ML for malware detection. Windows Portable Executable (PE) files are processed to extract features. It contains information like if an executable was packed, Application Programming Interface (API) calls used in the executable, networking-related API calls, and opcode. The ML models are trained and tested using extracted features to find the best-performing algorithm. The author proposed a three-dimensional taxonomy where they track the objective of the analysis, the feature, and the feature extraction method. A challenge to ML-based malware detection is anti-analysis techniques like packing, API hashing, and Virtual Machine (VM) and debugger detection.

In the article by Dasgupta et al. (2022). Machine learning in cybersecurity: a comprehensive survey, noted that the ML algorithms are prone to attacks during the training and testing phase. This vulnerability can impact the accuracy rate of the ML model. If the ML model is not trained and tested with quality training and testing data sets, then it can impact the accuracy of the ML model.

Bharadiya (2023). Machine Learning in Cybersecurity: Techniques and Challenges, identified additional challenges with implementing ML models in cybersecurity. ML models are vulnerable to data poisoning, meaning if the dataset elements are mislabeled, then the ML model can generate False Positive (FP) and False Negative (FN) results, impacting the accuracy of the ML model. The author also mentions evasion attacks where minor variations in test data can mislead the ML algorithm and impact the performance. Since the accuracy of the ML model depends on the quality of training data, it is challenging to find accurate real-world cyber-attack data. In an imbalanced data set where the amount of True Positive (TP) and True Negative (TN) data is disproportionate, the ML model can produce biased detection, causing poor performance of the ML model. The ML models in the cybersecurity domain need to process a large amount of data in real-time and produce accurate results. It requires a scalable number of resources, which can be challenging and costly. However, with the emergence of Cloud-based computing, this challenge was solved. Training ML models may also require sensitive, protected, and private data in some applications. With increasing regulations and privacy laws like the General Data Protection Regulation (GDPR), it gets challenging to get the data owner's consent to collect and process the data Sahingoz et al. (2019).

4. CONCLUSION

Based on the review, we have identified the use of machine learning in the cyber security domain to implement IDS, NIDS, malware detection, spam, and phishing detection. Building ML-based cybersecurity products can be challenging as their performance depends on the quality and availability of training and testing data.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Bharadiya, J. (2023). Machine Learning in Cybersecurity: Techniques and Challenges. European Journal of Technology, 7(2), 1-14. https://doi.org/10.47672/ejt.1486

Dasgupta, D., Akhtar, Z., & Sen, S. (2022). Machine Learning in Cybersecurity: A Comprehensive Survey. The Journal of Defense Modeling and Simulation, 19(1), 57-106. https://doi.org/10.1177/1548512920951275

Rao, B. B., & Swathi, K. (2017). Fast kNN Classifiers for Network Intrusion Detection System. Indian J. Sci. Technol., 10(14), 1-10. https://doi.org/10.17485/ijst/2017/v10i14/93690

Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine Learning Based Phishing Detection from URLs. Expert Systems with Applications, 117, 345-357. https://doi.org/10.1016/j.eswa.2018.09.029

Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of Machine Learning Techniques for Malware Analysis. Computers & Security, 81, 123-147. https://doi.org/10.1016/j.cose.2018.11.001

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., & Wang, C. (2018). Machine Learning and Deep Learning Methods for Cybersecurity. Ieee Access, 6, 35365-35381. https://doi.org/10.1109/ACCESS.2018.2836950