Cybersecurity for Data Science: Issues, Opportunities, and Challenges

Mamoona Humayun, N. Z. Jhanjhi, M. N. Talib, Mudassar Hussain Shah, and G. Suseendran

Abstract Cybersecurity (CS) is one of the critical concerns in today's fast-paced and interconnected world. Advancement in IoT and other computing technologies had made human life and business easy on one hand, while many security breaches are reported daily. These security breaches cost millions of dollars loss for individuals as well as organizations. Various datasets for cybersecurity are available on the Internet. There is a need to benefit from these datasets by extracting useful information from them to improve cybersecurity. The combination of data science (DS) and machine learning (ML) techniques can improve cybersecurity as machine learning techniques help extract useful information from raw data. In this paper, we have combined DS and ML for improving cybersecurity. We will use the CS dataset, and ML techniques will be applied to these datasets to identify the issues, opportunities, and cybersecurity challenges. As a contribution to research, we have provided a framework that will provide insight into ML and DS's use for protecting cyberspace from CS attacks.

Keywords Cybersecurity · Machine learning · Data science · IoT · Attacks

M. Humayun (✉)
College of Computer and Information Sciences, Jouf University, Al-Jouf, Saudi Arabia
e-mail: mahumayun@ju.edu.sa

N. Z. Jhanjhi
School of Computer Science and Engineering (SCE), Taylor's University, Selangor, Malaysia

M. N. Talib
Papua New Guinea University of Technology, Lae, Papua New Guinea
e-mail: muhammad.talib@pnguot.ac.pg

M. H. Shah
Department of Communication and Media Studies, University of Sargodha, Sargodha, Pakistan

G. Suseendran
Department of Information Technology, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

S.-L. Peng et al. (eds.), *Intelligent Computing and Innovation on Data Science*,
Lecture Notes in Networks and Systems 248,
https://doi.org/10.1007/978-981-16-3153-5_46

5 Conclusion and Future Works

This paper provides deeper insights into three well-known disciplines: DS, ML, CS, and interrelationship. According to our findings, DS and ML help make cybersecurity decisions. Various datasets exist related to different security attacks in multiple domains; these datasets can be used for future attack prediction. To extract intelligent security solutions from existing security datasets, ML and DS techniques are helpful. To synthesize the obtained information and contribute, we have developed a framework based on the interrelationship between CS, DS, and ML. The proposed framework will help security practitioners get intelligent cybersecurity solutions.

In the future, we are planning to apply the proposed framework to a real security dataset for extracting intelligent security solutions.

References

1. Von Solms R, Van Niekerk J (2013) From information security to cyber security. Comput Secur 38:97–102
2. Humayun M et al (2020) Cyber security threats and vulnerabilities: a systematic mapping study. Arab J Sci Eng 1–19
3. Humayun M, Jhanjhi N, Alamri M (2020) IoT-based secure and energy efficient scheme for E-health applications. Indian J Sci Technol 13(28):2833–2848
4. Chiappetta A (2017) Hybrid ports: the role of IoT and cybersecurity in the next decade. J Sustain Dev Transp Logistics 2(2):47–56
5. Almusaylim ZA, Zaman N (2019) A review on smart home present state and challenges: linked to context-awareness internet of things (IoT). Wireless Netw 25(6):3193–3204
6. Tuor A et al (2017) Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. arXiv preprint arXiv:1710.00811
7. Foroughi F, Luksch P (2018) Data science methodology for cybersecurity projects. arXiv preprint arXiv:1803.04219
8. https://www.weforum.org/agenda/2020/09/rethinking-risk-management-and-compliance-age-of-ai-artificial-intelligence/
9. Sarker IH et al (2020) Cybersecurity data science: an overview from machine learning perspective. J Big Data 7(1):1–29
10. https://www.comparitech.com/vpn/cybersecurity-cyber-crime-statistics-facts-trends/
11. Galeano-Brajones J et al (2020) Detection and mitigation of DoS and DDoS attacks in IoT-based stateful SDN: an experimental approach. Sensors 20(3):816
12. Meidan Y et al (2018) N-baiot—network-based detection of iot botnet attacks using deep autoencoders. IEEE Pervasive Comput 17(3):12–22
13. https://analyticsindiamag.com/top-10-datasets-for-cybersecurity-projects/
14. Shahzadi S et al. (2021) Machine learning empowered security management and quality of service provision in SDN-NFV environment
15. Ahmad F et al. (2021) Prediction of COVID-19 cases using machine learning for effective public health management
16. Yener B, Gal T (2019) Cybersecurity in the era of data science: examining new adversarial models. IEEE Secur Priv 17(6):46–53

17. Wang T et al (2020) Implementation of a real-time psychosis risk detection and alerting system based on electronic health records using CogStack. JoVE (J Visualized Exp) 159: e60794
18. Almulhim M, Zaman N (2018) Proposing secure and lightweight authentication scheme for IoT based E-health applications. In: 2018 20th international conference on advanced communication technology (ICACT). IEEE, pp 481–487
19. Fraley JB, Cannady J (2017) The promise of machine learning in cybersecurity. In: SoutheastCon 2017. IEEE
20. https://searchsecurity.techtarget.com/tip/Unpack-the-use-of-AI-in-cybersecurity-plus-pros-and-cons