IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

A Review on Machine Learning Techniques for Cyber Security in the Last Decade

Ishita Bhadekar[1], Aarya S Dawre[2], Tanmay Chavhan[3], Vishakha N. Pawar[4]
MKSSS's College of Engineering for Women, Pune[1]
Guru Gobind Singh College Of Engineering And Research Centre, Nashik, Maharashtra, India[2,3]
Guru Gobind Singh Polytechnic, Nashik, Maharashtra, India[4]

Abstract: *Universal growth and usage of the Internet and mobile applications have elongated cyberspace. The cyberspace has become more exposed to automated and protracted cyberattacks. Cyber security approaches provide enrichments in security processes to detect and react against cyberattacks. The formerly used security systems are no longer enough since cybercriminals are sharp enough to avoid conventional security systems. Conventional security systems privation efficiency in perceiving formerly unseen and polymorphic security attacks. Machine learning (ML) techniques are playing a vigorous role in several applications of cyber security. However, regardless of the ongoing achievement, there are substantial challenges in safeguarding the honesty of Machine Learning systems. There are incentivized nasty opponents present in the cyberspace that are eager to game and adventure such Machine Learning vulnerabilities. This paper aims to afford a complete outline of the challenges that Machine Learning techniques expression in shielding cyberspace against attacks, by bestowing a literature on Machine Learning techniques for cyber security inclusive of intrusion detection, spam detection, and malware detection on computer networks and mobile networks in the last decade. It also provides brief descriptions of each ML method, frequently used security datasets, essential ML tools, and evaluation metrics to evaluate a classification model. It finally discusses the challenges of using ML techniques in cyber security. This paper provides the latest extensive bibliography and the current trends of ML in cyber security.*

Keywords: Deep Learning, Cyber Security, Malware, Intrusion Detection, Spam, Machine Learning

I. INTRODUCTION

The Internet is progressively fetching a widely developed source of both information and (online) services. There is speedy growth in Internet usage habit: in 2017, about 50% of the total world population used the Internet as a basis of information [1]. This number increased up to 80% in developed countries [2]. The key purpose of the Internet is to transfer data from one node to one more over the network. Internet is a widespread group of millions of discrete interconnected computers, networks, and associated devices. The modernisation of computer systems, networks, and mobile devices has powerfully increased the usage of the Internet. Therefore, the Internet has come to be the target of cybercriminals and enemies [3]. A protected and firm computer system must guarantee the confidentiality, availability, and integrity of information. The integrity and security of a computer system are cooperated when an illegal penetration, illegal individual or program pass in a computer or network planning to harm or disorder the normal flow of activities [4]. Cyber security is the set of security actions that can be taken to guard the cyberspace and user assets in contradiction of unauthorized access and attacks. The core objective of a cyber defence system is that data should be private, fundamental, and available [5].

National defence plays a vital role in the integrity of any country. Computer networks are intended to deliver controls, which permit solitary authorised persons to access data. Bush Administration started the Comprehensive National Cyber Security Initiative (CNCSI) in January 2008 [6]. The resolutions of the creativity were to highpoint numerous issues for incidence identification of present and developing cyber security threats, discovery and plugging existing cyber vulnerabilities, and capturing actors that were demanding to gain access to secure central information systems. The cyberattack that should be highlighted is the attack that suffered by Estonia in 2007. Different Estonian financial,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

educational, and newspaper websites were hacked for three weeks [8]. It was considered the first cyberwar, which took the attention of the NATO Bucharest Summit Declaration. NATO announced a policy on cyber defence in 2008 [9].

Inherent and internal weakness in the configuration and implementation of a computer system and network creates vulnerabilities that render them susceptible to cyberattacks and threats. Incorrect configuration, lack of adequate procedures, inexperienced or untrained personnel are examples of vulnerabilities in building a computer network system. These vulnerabilities increase the chances of threats and attacks within a network or from outside a network. A significant number of people from different fields are becoming dependent on cyber networks. Using a particular penetration technique, an agent that causes harmful and undesirable effects in activities and behaviour of a computer or network is called a threat [10]. Cyber security is to protect the integrity of the data, networks, and programs from cyber threats to cyberspace [11].

Since the initiation of the rst PC infection in 1970, there is a race among cybercriminals and safeguards [12]. It is getting increasingly testing to fight against these network protection assaults and to keep a coordinate with the speed of safety assaults. At present, specialists are zeroing in on the critical need of finding new robotized security techniques to adapt to these security challenges. Truly outstanding and powerful considered practice is to utilize robotized machine learning strategies to distinguish new and beforehand inconspicuous digital dangers [13].

A. DEVELOPMENT OF MACHINE LEARNING AND CYBER SECURITY IN LAST DECADE

The utilization of AI and ML strategies is getting extended quickly in various regions of life, for example, finance [14]-[16], instruction [17], medication[18]-[21], fabricating industry [22], and especially in the field of digital protection [23]-[28]. ML methods are assuming a crucial part in various uses of the digital protection for early recognition what's more expectation of various assaults like spam classification [29]-[32], misrepresentation identification [33]-[36], malware identification [37]-[40], phishing [41]-[43], darkweb or deepweb destinations [44], [45], and interruption identification [46][49]. ML strategies can address the shortage accessible of required work force with mastery in these specialty cybercrime discovery advances. In addition, vivacious methodologies are required to recognize and respond against the cyberattacks of the new age (computerized and transformative). AI is one of the potential answers for act rapidly against such assaults since ML can gain from encounters and react to more up to date assaults on schedule. There is a ton of writing accessible on the Internet that portrays the utilization of ML for the predication of digital dangers on darkweb or deepweb.

Mohammad et al. [45] applied ML models to anticipate digital dangers by assessing the informal organizations of programmers on darkweb.

Sarkar et al. [50] utilized a set-up of informal organization highlights what's more applied ML models to foresee whether there would be an assault on a specific association on the anticipated date or on the other hand not. They have performed tests by social affair the information from 53 gatherings on darkweb. The predications of assaults through the conversation of darkweb are out of degree from this review paper. Nonetheless, late headways around here can be found in [51][54]. Figure 1 portrays the patterns of network safety and the two regions connected with information science (i.e. ML and profound learning (DL)) overall and independently.

We had got the details from Scopus on June 23, 2020 profound learning can be considered as a subset of machine learning, a few articles have utilized the term of profound learning rather than AI in managing digital protection. We have looked and really looked at the patterns of digital protection what's more ML and the patterns of digital protection and DL independently to concentrate on them in more subtleties. We have shown the patterns in Figure 1 throughout the previous decade. In the first half of the ten years, the ML models were applied for the identification of assaults on cloud security, malware, and interruptions. Notwithstanding, the pattern has been expanded at a marvellous rate with the arising improvement in the field of profound learning.

B. COMMITTMENT OF THE PAPER

The reason for this article is to audit the critical machine learning procedures applied in network protection and point out the pattern of involving AI procedures for digital protection.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

We have given a concise portrayal of AI strategies, and how AI procedures have been, or on the other hand could be, utilized to distinguish and order cyberattacks, for example, interruption identification, malware recognition, and spam discovery on both PC organizations and mobiles or cell phones gadgets.

Any hunt system should permit the culmination of the search to be surveyed. To recognize significant commitments in network protection and AI, IEEE Xplore, ACM computerized library, Emerald Insight, SpringerLink and ScienceDirect were questioned for papers having ('Machine Learning' furthermore 'Network protection'), ('Machine Learning' and 'Network protection'), ('Profound Learning' and 'Network protection'), ('Deep Learning' and 'Network protection'), ('Machine Learning' and 'Malware'), ('Machine Learning' and 'Interruption Detection'), ('AI' and 'Spam'), ('Deep Learning' and 'Malware'), ('Deep Learning' and 'Interruption Detection'), and ('Profound Learning' and 'Spam') in title, dynamic or watchwords. Likewise, Web of Science, Google Scholar, and Scopus were questioned to twofold actually look at the findings and to find other related papers in less-known libraries. Google Scholar was moreover utilized for forward and in reverse hunts. We have centered on late headways over the most recent decade. These on the web information bases were picked as they offer the most significant peer-audited full-text diaries and meeting procedures, book sections, and reports covering the field of machine learning and digital protection. Altogether, 7915 records were recovered. The copied things were eliminated. The title also dynamic of 1728 archives were screened to distinguish possible articles. The full-text appraisal of 770 was made as indicated by the significance of the incorporation measures. Further, 486 examinations were excluded. We have avoided the articles that were examining (1) informal community legal sciences, (2) immaterial digital dangers, (3) dangers to digital actual lattices, (4) dangers to cloud security, IoT gadgets, (5) savvy matrices, and brilliant urban areas, furthermore (6) satellite correspondence, 5G and remote correspondence.

With forward and in reverse hunt, 28 additional investigations were recovered. Altogether, 312 investigations were finally chosen for information extraction reason. Figure 2 shows the course of article incorporation and determination. Likewise, the past study and audit articles were utilized to give a thorough review of AI methods in digital security.

It is expected that the used search terms will cover most, if not all, of the work incorporating machine learning methods for cyber security. Nevertheless, Google Scholar is further utilized to check the citation of found papers (forward-searching) to update our search and to look for other scientific resources to make sure nothing is neglected. The last update of the searching of papers was done on May 3rd, 2020. Table 1 depicts the list of acronyms used in this article for convenient referencing. We are unaware of any existing survey that provides the application of ML techniques in cyber security on both computer and mobile networks. Our work also presented commonly used ML tools, security datasets, graphical summary of significant components of cyber security and available ML techniques to fight against threats and attacks on cyberspace, and future challenges such as trustworthiness and adversarial machine learning under one umbrella. Table 2 presents a comparison of our paper with existing surveys and review articles. Many current surveys, either present applications in a particular domain or lack of giving basic knowledge that a new researcher requires to get in or understand this domain. Furthermore, most of the survey articles discuss particular threats and attacks on a network only. We have focused on significant cyber security such as intrusion detection, malware detection, and spam classification on both networked computers and mobile devices. In particular, machine learning techniques have not only increased threats on computer networks but also held a lot of promises for detection and classification of attacks and threats on mobile devices and networks. Our survey covers cyber threats on both mobile devices and computer networks. Comparing to existed survey papers in the area, our survey is inclusive and unique in providing the following aspects: providing basic insights of cyber security threats on both mobile devices and computer networks, giving descriptions of commonly used security datasets, summarizing the stateof-the-art ML techniques to handle these threats, indicating popular ML tools, describing evaluation metrics to evaluate the performance of ML techniques, and pointing out current challenges of ML techniques for cyber security. We have provided a graphical summary of major components of cyber security and available machine learning techniques to fight against these attacks on cyberspace. The last updating on the paper's citations count (source: Google Scholar) was done on
June 05, 2020, in Table 2.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

CONCLUSION

Digital protection has turned into an issue of concern universally in accomplishing upgrades in safety efforts to recognize and respond against cyberattacks. The recently utilized ordinary security frameworks are no longer sufficient in light of the fact that those frameworks need efficiency in identifying already inconspicuous and polymorphic assaults. AI procedures are playing an imperative job in various uses of network safety frameworks.

Our audit here has uncovered a quickly developing interest in AI and network protection in the scholastics and industry, which has brought about a developing number of distributions, especially somewhat recently. In this paper, we have overcome any barrier between ML procedures and dangers to PC organizations and versatile correspondence by introducing an extensive review of the hybrids between the two regions. This overview presents the writing audit on machine learning methods for interruption discovery, spam recognition, also malware discovery on PC organizations and portable gadget somewhat recently.

This paper briefly presents the utilizations of machine learning models in the field of network protection, primarily on the headway of the most recent decade. There are idiosyncrasies of each digital danger that make it difficult in any event, for the cutting edge ML model in managing such cyberattacks. It is difficult to give one proposal to all the assaults, in view of one model. Different rules like location rate, time intricacy, and classification time to distinguish new what's more zero-day assaults, and precision of a ML model ought to be considered while choosing a specific model to distinguish a cyberattack. We have depicted the nuts and bolts of network safety for example, the classification of cyberattacks on cell phone what's more PC organizations. Because of the significance of ML, we have likewise depicted the underpinnings of AI, subtypes, and significant methods for a fledgling to get a better understanding into this area. We know nothing about any work that examines the utilizations of ML methods in network safety area both on cell phone and PC networks in one paper. We have portrayed a graphical outline of the assaults threatened to the internet and existing ML methods to fight against these cybercrimes. We have introduced an outline of a few famous ML instruments. We have likewise given the assessment measurements to assess the working of any classifier.

Dataset is exceptionally essential for the preparation and testing of ML models. We have introduced a depiction of ordinarily utilized security datasets. There is the inaccessibility of delegate furthermore benchmark datasets for every danger space. Machine learning strategies were not principally intended to work with network safety. Avoidance can without much of a stretch moron the ML model by giving antagonistic information sources. Reliable AI is the protected utilization of AI methods for the internet to give some significant level rightness ensures rather than speed and precision of the model. We have likewise briey summed up a portion of the significant difficulties of utilizing machine learning methods in digital protection as well as given a broad book index around here. The referenced difficulties genuinely deserve consideration for future examination.

REFERENCES

[1] ICT Fact and Figures 2017. Accessed: Jun. 1, 2020. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf

[2] ICT Facts and Figures, International Telecommunication Union. (2017). Telecommunication Development Bureau. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx (accessed Oct. 09, 2019).

[3] D. K. Bhattacharyya and J. K. Kalita, Network Anomaly Detection: A Machine Learning Perspective. London, U.K.: Chapman & Hall, 2013.

[4] V. Ambalavanan, ``Cyber threats detection and mitigation using machine learning,'' in Handbook of Research on Machine and Deep Learning Applications for Cyber Security. Hershey, PA, USA: IGI Global, 2020, pp. 132149.

[5] T. Thomas, A. P. Vijayaraghavan, and S. Emmanuel, ``Machine learning and cybersecurity,'' in Machine Learning Approaches in Cyber Security Analytics. Singapore: Springer, 2020, pp. 3747. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-1706-8_3

[6] The Comprehensive National Cybersecurity Initiative. Accessed: Jun. 1, 2020. [Online]. Available: https://obamawhitehouse. archives.gov/issues/foreign-policy/cybersecurity/national-initiative

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

[7] The White House, Remarks by APHSCT Lisa O. Monaco at the International Conference on Cyber Security. Accessed: Oct. 17, 2019. [Online]. Available: https://obamawhitehouse.archives.gov/thepress-office/2016/07/26/remarks-aphsct-lisa-o-monaco-internationalconference-cyber-security
[8] 10 Years After the Landmark Attack on Estonia, Is the World Better Prepared for Cyber Threats? Accessed: Jun. 1, 2020. [Online]. Available: https://foreignpolicy.com/2017/04/27/10-years-after-the-landmarkattack-on-estonia-is-the-world-better-prepared-for-cyber-threats/
[9] North Atlantic Treaty Organization. (Apr. 3, 2008). Bucharest Sum-mit Declaration. Issued by the Heads of State and Government Participating in the Meeting of the North Atlantic Council in Bucharest. Accessed: Oct. 9, 2019. [Online]. Available: https://www.nato.int/cps/en/natolive/official_texts_8443.htm
[10] F. Farahmand, S. B. Navathe, P. H. Enslow, and G. P. Sharp, ``Managing vulnerabilities of information systems to security incidents," in Proc. 5thInt. Conf. Electron. Commerce (ICEC), 2003, pp. 348354.
[11] D. Craigen, N. Diakun-Thibault, and R. Purse, ``Defining cybersecurity," Technol. Innov. Manage. Rev., vol. 4, no. 10, pp. 1321, Oct. 2014.
[12] P. Szor, The Art of Computer Virus Research and Defense: ART COMP VIRUS RES DEFENSE_p1. London, U.K.: Pearson, 2005.
[13] I. Firdausi, C. Lim, A. Erwin, and A. S. Nugroho, ``Analysis of machine learning techniques used in behavior-based malware detection," in Proc. 2nd Int. Conf. Adv. Comput., Control, Telecommun. Technol., Dec. 2010, pp. 201203.
[14] S. Gu, B. T. Kelly, and D. Xiu, ``Empirical asset pricing via machine learning," in Proc. 31st Australas. Finance Banking Conf., Chicago Booth Res. Paper 18-04, Yale ICFWorking Paper 2018-09, Sep. 2019. [Online]. Available: https://ssrn.com/abstract=3159577
[15] P. Mathur, ``Overview of machine learning in finance," in Machine Learning Applications Using Python. Berkeley, CA, USA: Apress, 2019, pp. 259270. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4842-3787-8_13
[16] S. Emerson, R.Kennedy, L. O'Shea, and J. O'Brien, ``Trends and applications of machine learning in quantitative finance," in Proc. 8th Int. Conf. Econ. Finance Res. (ICEFR), 2019, pp. 19.
[17] K. Shaukat et al., ``Student's performance in the context of data mining," in Proc. 19th Int. Multi-Topic Conf. (INMIC), 2016.
[18] S. Jha and E. J. Topol, ``Adapting to artificial intelligence: Radiologists and pathologists as information specialists," Jama, vol. 316, no. 22, pp. 23532354, 2016.
[19] A. I. Tekkesin, ``Articial intelligence in healthcare: Past, present and future," Anatolian J. Cardiol., vol. 2, no. 4, pp. 230243, 2019.
[20] K. Shaukat, N. Masood, A. Bin Shafaat, K. Jabbar, H. Shabbir, and S. Shabbir, ``Dengue fever in perspective of clustering algorithms," 2015, arXiv:1511.07353. [Online]. Available: http://arxiv.org/abs/1511.07353
[21] K. S. Dar and S. M. U. Azmeen, ``Dengue fever prediction: A data mining problem," J. Data Mining Genomics Proteomics, vol. 6, no. 3, pp. 15, 2015.
[22] B.-H. Li, B.-C. Hou,W.-T. Yu, X.-B. Lu, and C.-W. Yang, ``Applications of artificial intelligence in intelligent manufacturing:Areview," Frontiers Inf. Technol. Electron. Eng., vol. 18, no. 1, pp. 8696, 2017.
[23] C. Virmani, T. Choudhary, A. Pillai, and M. Rani, ``Applications of machine learning in cyber security," in Handbook of Research on Machine and Deep Learning Applications for Cyber Security. Hershey, PA, USA: IGI Global, 2020, pp. 83103.
[24] R. Calderon, ``The benets of artificial intelligence in cybersecurity," La Salle Univ., Philadelphia, PA, USA, Tech. Rep. Winter 1-15-2019, 2019.