Blockchain Frontier Technology (B-Front)
Vol. 3 No. 1 July 2023

P-ISSN: 2808-0831
E-ISSN: 2808-0009

Learning Cyber Security and Machine Engineering at the University

Rickho Rizky Karuniawan[1], Sugeng Santoso[2], Muhamad Al Fikri[3], M. Argadilah[4], Wisnu Ardi Pamungkas[5]
Management Information System, University of Raharja[1,3,4]
Department of Information Technology, University of Raharja[2]
Computer System, University of Raharja[5]
Indonesia
e-mail: rickho@raharja.info, sugeng.santoso@raharja.info, alfikri@raharja.info, m.argadilah@raharja.info, wisnu.ardi@raharja.info

Author Notification
28 Desember 2022
Final Revised
17 June 2023
Published
27 June 2023

Karuniawan, R. R., Santoso, S., Fikri, M. A., Argadilah, M., & Pamungkas, W. A. (2022). Learning Cyber Security and Machine Engineering at the University. Blockchain Frontier Technology, 3(1), 89–94.
DOI: https://doi.org/10.34306/bfront.v3i1.242

Abstract

In this review, key literature reviews on network analysis and intrusion detection using machine learning (ML) and deep learning (DL) approaches are explained. It also provides a brief lesson description for each ML/DL procedure. This paper covers the datasets used in machine learning techniques, which are the main instruments for evaluating network traffic and detecting irregularities. Data holds a key role in ML/DL approaches. We also go into further detail about the problems with using ML/DL for cybersecurity and make suggestions for future research.

Keywords: *Machine learning, Cyber Security, Machine Engineering, Learning*

1. Introduction

The applications of cyber security are considered in this work, which also includes the findings of a literature review on machine learning (ML), deep learning (DL), and data mining (DM) techniques [1]. It also describes (ML/DL)/DM approaches and their applicability to problems with cyber intrusion detection. The study studies the difficulty of several (ML/DL)/DM algorithms in addition to offering a set of comparison criteria for (ML/DL)/DM approaches [2]. Best practices and a list of suggestions are then provided in consideration of the characteristics of the cyber situation [3]. Cybersecurity is the umbrella term for a collection of procedures and tools designed to guard against unauthorized access to, hacking into, altering, or destroying networks, computers, data, and software [4]. Cyber security systems primarily consist of network security systems and computer (host) security systems. Each of these entities must contain at least one intrusion detection system (IDS), firewall, and antivirus program [5]. IDSs do not only identify information systems; they also find and ascertain their unauthorized usage, change, duplication, and damage. Security breaches include both external and internal intrusions (intra-organization assaults and inter-organizational attacks) [6]. The surge in the frequency of cyberattacks has made artificial intelligence- and machine-learning-based approaches for detecting security risks an integral part of our daily life. The best security apps are provided with consideration for their implementation and security requirements thanks in large part to the security standards [7]. From this angle, they are essential for cyber security research, such as intrusion detection systems. While certain datasets and machine and deep learning algorithms can be supported by research studies, there is little information in the literature on (ML/DL)/DM approaches for cyber security and datasets linked to security [8]. This survey article focuses on the (ML/DL)/DM approaches with

an emphasis on the (ML/DL)/DM methods and their descriptions with an emphasis on cyber security. Several reviews as well as other studies using these techniques have also been published [9]. In contrast to past evaluations, the focus of our paper will be on publications that meet specific criteria. Google Scholar searches were carried out using methods including "machine learning" or "deep learning," cybersecurity, and "data mining [10]." One of the key issues was highly cited papers using well-known approaches. However, it was realized that fresh and inventive approaches might be disregarded, so a couple of these pieces were also chosen [11]. Most often, the decision to include at least one and preferably a few sample papers about each of the (ML/DL)/DM categories was made while choosing the papers to be included [12]. The rest of this essay is structured as follows: Sect. 2 provided a summary of some pertinent academic writing. The many (ML/DL)/DM techniques utilized in cyber security are described in Section 4. The topic of cybersecurity datasets for (ML/DL)/DM is covered in Section 4, and the paper is concluded with a quick overview of its main ideas and some concluding notes in Section 5 [13].

2. Literature Review

Machine learning applications offer a unique method for resolving basic scientific and technical puzzles with the use of computer software [14]. In the last 20 years, there have been significant advances in machine learning that are accessible to beginners. Machine learning was essentially started in a "black-box" laboratory environment, developed into a usable application, and is now being gradually used on a broad scale by commercial enterprises [15]. The advances of machine learning can be seen in the software programs for computer vision, natural language processing, speech recognition, robot control, and other new applications [16]. Major businesses like Amazon, Facebook, and Google employ machine learning to enhance the customer experience, advertise special offers, and suggest purchases [17]. It is far simpler for machine learning developers to train a system than it is to program the conventional input-process-output system [18]. They do this by creating simulations of expected results. Many different businesses have seen how machine learning has affected data-intensive problems like cybersecurity [19]. To bring about amazing changes, other subjects like biology, cosmology, and social science can also benefit from using machine learning. Machine learning can be used in novel ways to analyze and analyze experimental data. Theoretically, by working on machine learning algorithms, we can comprehend the ideas of "big data" better. Additionally, these technologies can be used in vertical applications to enhance associated performance measures [20].

In terms of special functionalities, machine learning algorithms might vary significantly (e.g., logistic regression, linear Regression, Naive Bayes, decision trees, random forest, support vector machines, deep learning, and Gradient Boosting algorithms) [21]. To produce an evolutionary method, machine learning presents innovative approaches to examine vast amounts of data. The subsequent generations of algorithms can also offer better optimization. The optimum way to process the vast amount of data is through machine learning and cybersecurity [22]. The networks and platforms are open to intrusion. The quantity of instruments used to scan and assess targets affects how effective these attacks are. The opponents utilize machine learning to intensify their attacks even more [23]. A few surveys on the security perspective of artificial intelligence and machine learning have recently been published in some journals [24]. In addition, emphasis was placed on recent literature on machine learning techniques utilized in the Internet of Things and intrusion detection for computer network security. After that, major literature reviews on machine learning (ML) and deep learning (DL) methods were explained for network analysis of intrusion detection along with an explanation of each ML/DL method. A thorough literature analysis on defensive tactics and security risks was offered during training and testing or inferring of machine learning from a data-driven view. Additionally, information on security concerns with artificial intelligence, particularly about the reinforcement and supervised learning algorithms, was provided. In

P-ISSN: 2808-0831
E-ISSN: 2808-0009

contrast, the life cycle of a machine learning-based system from training to inference was examined. This included updates on security risks and protective strategies [25].

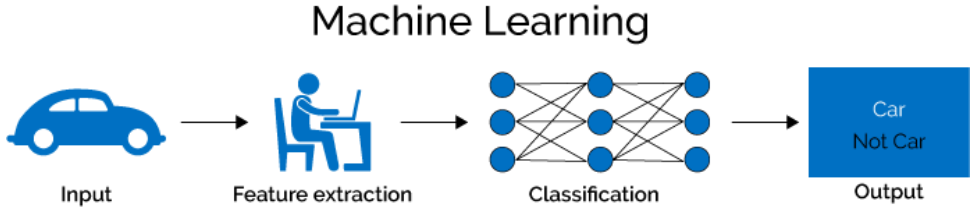
3. Cybersecurity Classification of Machine Learning Algorithms

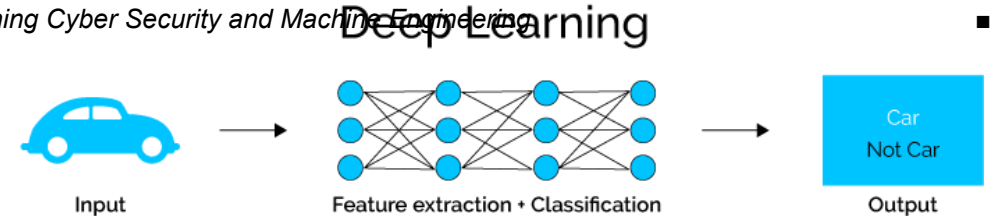
3.1 Machine Engineering Techniques

Deep learning, a novel area of machine learning, has recently received extensive attention and has been applied to intrusion detection. The results of the studies also show that deep learning outperforms conventional approaches. The authors have used a deep learning method based on a deep neural network for flow-based anomaly identification. The results of the experiment show that deep learning can also be used for anomaly identification in software-defined networks. In a network intrusion detection system, a deep learning-based technique is suggested employing self-taught learning (STL) on the benchmark NSL-KDD dataset. When compared to the methods outlined in earlier studies, the methodology is proven to perform better. However, this group of references emphasizes deep learning's capacity for feature reduction. The categorization is implemented using the conventional supervision model, while pre-training is mostly accomplished using deep learning techniques. It is an unusual task to use deep learning to perform classification directly, and there isn't any research in the literature that shows how well it performs in multiclass classification. RNNs are thought to be smaller neural networks. In this work, a three-layer RNN architecture with 41 features as inputs and four incursion categories as outputs was recommended for misuse-based IDs. The decreased RNNs lack deep learning's capacity for modeling high dimensional characteristics, and the layer nodes are only partially connected. The authors do not examine the model's performance in binary classification. learned how to construct an intrusion detection system using deep learning, and recurrent neural networks were suggested as a deep learning strategy for intrusion detection (RNN-IDS). Additionally, the model's effectiveness in binary classification and multiclass classification was examined. It was found that the number of neurons and various learning rates had an impact on how well the suggested model performed. According to the results of the experiments, the RNNIDS is particularly suitable for constructing a classification model with high accuracy. It displays higher performance than the conventional machine learning classification approaches in both multiclass and binary classification. The RNN-IDS model provides a fresh approach to the field of intrusion detection while also improving its efficacy.

3.2 Traditional Machine Learning Methods

The main goal of ML is the development of models that take input data and combine it with statistical analysis to predict an output value within a specified range. One of the fields of computer science that is fast expanding and has a wide range of applications is machine learning. ML algorithms can be divided into supervised, unsupervised, and reinforcement learning categories. Supervised algorithms are the well-known processes utilized in machine learning algorithms. Further divisions of supervised algorithms include regression and classification. In literature, a variety of machine learning algorithms are employed. Among the most popular machine-learning algorithms include Logistic Regression, Decision Tree, Naive Bayes, SVM, K-Means, KNN, and Random Forest.





Source: Towards Data Science

Figure 1. Traditional Machine Learning Methods

4. Digital Security Sets

Various research organizations are now preparing data for both their study and release to community repositories. This section explains the current security-related datasets using machine learning and artificial intelligence research.

4.1 UNSW-NB15 Dataset

The UNSW-NB15 Dataset is said to have been produced by the IXIA Perfect Storm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS). This dataset contains about an hour's worth of anonymized traffic traces from a DDoS assault that occurred in 2007. Fuzzers, Backdoors, Analysis, Exploits, DoS, Reconnaissance, Generic, Worms, and Shellcode are among the nine key attack types represented in this dataset.

4.2 KDD Cup 1999 Dataset

In addition to KDD Cup 1999, the material retrieved from MIT Lincoln Labs also includes tcpdump and BSM list files. This particular dataset included information gathered as part of the DARPA 98 IDS evaluation program. This dataset is also regarded as benchmark data for the evaluation of intrusion detection systems. The KDD '99 is one of the most well-known data sets for evaluating the effectiveness of anomaly detection techniques. Currently, several researchers are using the KDD dataset.

4.3 CTU-13 (Czech Technical University) Dataset

The CTU-13 (Czech Technical University) dataset is a collection of malware seizures from 13 different families in a real-world network environment. The goal of this dataset is to gather actual mixed botnet traffic. Keeping the aforementioned in mind, regular traffic is produced by confirmed regular hosts, whereas botnet traffic is produced by infected hosts. One advantage of using this dataset is that it is a thoroughly annotated dataset that captures the procedures carried out in a controlled environment.

4.4 ISOT (Information Security and Object Technology) Dataset

The ISOT dataset is made up of publicly accessible botnets and regular datasets, totaling 1,675,424 traffic flows. Regarding the malicious traffic in ISOT, it was obtained through the French honeynet operation, which included the Storm and Waledac botnets.

4.5 HTTP CSIC 2010 Dataset

The HTTP CSIC 2010 dataset is a collection of thousands of online requests that were generated automatically and developed at the Information Security Institute of CSIC (Spanish Research National Council). This dataset can be used to evaluate web attack defense systems. This data is compiled from over 25,000 aberrant queries and almost 6,000 typical

requests. Additionally, HTTP queries are classified as abnormal or regular. With the help of this dataset, web detection was effectively done.

5. Conclusion and Future Work

The protection of computer systems from cyber-attacks is one of the most important challenges for both national and international security. This article uses a variety of machine learning approaches to give a survey on security issues. Several datasets have been used for a variety of research projects. Additionally, machine learning and artificial intelligence play a big part in the defense of computer systems. This study explains the literature review of ML/DL and DM approaches used in cyber. We carefully searched for example articles describing various ML/DL and DM techniques in the cyber realm and determined common classes of varied datasets along with their benefits and drawbacks. In the future, we intend to produce more datasets that will be accessible to everyone.

References

[1] N. Sari, W. A. Gunawan, P. K. Sari, I. Zikri, and A. Syahputra, "Analisis Algoritma Bubble Sort Secara Ascending Dan Descending Serta Implementasinya Dengan Menggunakan Bahasa Pemrograman Java," *ADI Bisnis Digit. Interdisiplin J.*, vol. 3, no. 1, pp. 16–23, 2022.

[2] M. Azmi, M. S. Shihab, D. Rustiana, and D. P. Lazirkha, "The Effect Of Advertising, Sales Promotion, And Brand Image On Repurchasing Intention (Study On Shopee Users)," *IAIC Trans. Sustain. Digit. Innov.*, vol. 3, no. 2, pp. 76–85, 2022.

[3] D. Apriani, T. Ramadhan, and E. Astriyani, "Kerja Lapangan Berbasis Website Untuk Sistem Informasi Manajemen Praktek (Studi Sistem Informasi Program Studi Kasus Merdeka Belajar Kampus Merdeka (MBKM) Universitas Raharja," *ADI Bisnis Digit. Interdisiplin J.*, vol. 3, no. 1, pp. 24–29, 2022.

[4] D. Rustiana, J. D. Pratama, T. Mudabbir, and M. A. Fahmi, "Adoption Computerized Certificate Transparency And Confidentiality," *Int. J. Cyber IT Serv. Manag.*, vol. 2, no. 1, pp. 1–10, 2022.

[5] P. A. Sunarya, F. Andriyani, Henderi, and U. Rahardja, "Algorithm automaticPrawira, M., Sukmana, H. T., Amrizal, V., & Rahardja, U. (2019). A Prototype of Android-Based Emergency Management Application. 2019 7th International Conference on Cyber and IT Service Management, CITSM 2019. https://doi.org/10.1109/CI," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.5 Special Issue, pp. 387–391, 2019, doi: 10.30534/ijatcse/2019/6281.52019.

[6] P. Hendriyati, F. Agustin, U. Rahardja, and T. Ramadhan, "Management Information Systems on Integrated Student and Lecturer Data," *APTISI Trans. Manag.*, vol. 6, no. 1, pp. 1–9, 2022.

[7] U. Rahardja, "Meningkatkan Kualitas Sumber Daya Manusia Dengan Sistem Pengembangan Fundamental Agile," *ADI Bisnis Digit. Interdisiplin J.*, vol. 3, no. 1, pp. 63–68, 2022.

[8] R. Widayanti, U. Rahardja, F. P. Oganda, M. Hardini, and V. T. Devana, "Students Formative Assessment Framework (Faus) Using the Blockchain," in *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, 2021, pp. 1–6.

[9] U. Rahardja, Q. Aini, F. P. Oganda, and V. T. Devana, "Secure Framework Based on Blockchain for E-Learning During COVID-19," in *2021 9th International Conference on Cyber and IT Service Management (CITSM)*, 2021, pp. 1–7.

[10] F. P. Oganda, M. Hardini, and T. Ramadhan, "Pengaruh Penggunaan kontrak cerdas pada Cyberpreneurship Sebagai Media Pemasaran dalam Dunia Bisnis," *ADI Bisnis Digit. Interdisiplin J.*, vol. 2, no. 1, pp. 55–64, 2021.

[11] S. Azizah and B. P. K. Bintoro, "Determining Factors of Continuance Intention to Use QR Code Mobile Payment on Urban Millennials in Indonesia Empirical Study on Mobile Payment Funds," *ADI J. Recent Innov.*, vol. 3, no. 2, pp. 121–138, 2022.

[12] T. C. Husnadi, T. Marianti, and T. Ramadhan, "Determination of shareholders' welfare with financing quality as a moderating variable," *APTISI Trans. Manag.*, vol. 6, no. 2, pp. 191–208, 2022.

[13] D. Rifai, S. Fitri, and I. N. Ramadhan, "Perkembangan Ekonomi Digital Mengenai Perilaku Pengguna Media Sosial Dalam Melakukan Transaksi," *ADI Bisnis Digit. Interdisiplin J.*, vol. 3,

no. 1, pp. 49–52, 2022.
[14] B. Rawat, N. Mehra, A. S. Bist, M. Yusup, and Y. P. A. Sanjaya, “Quantum Computing and AI: Impacts & Possibilities,” *ADI J. Recent Innov.*, vol. 3, no. 2, pp. 202–207, 2022.
[15] E. S. Pramono, D. Rudianto, F. Siboro, M. P. A. Baqi, and D. Julianingsih, “Analysis Investor Index Indonesia with Capital Asset Pricing Model (CAPM),” *Aptisi Trans. Technopreneursh.*, vol. 4, no. 1, pp. 36–47, 2022.
[16] E. A. Nabila, S. Santoso, Y. Muhtadi, and B. Tjahjono, “Artificial Intelligence Robots And Revolutionizing Society In Terms Of Technology, Innovation, Work And Power,” *IAIC Trans. Sustain. Digit. Innov.*, vol. 3, no. 1, pp. 46–52, 2021.
[17] Y. Durachman, A. S. Bein, E. P. Harahap, T. Ramadhan, and F. P. Oganda, “Technological and Islamic environments: Selection from Literature Review Resources,” *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 1, pp. 37–47, 2021.
[18] M. Saraswati, N. Lutfiani, and T. Ramadhan, “Kolaborasi Integrasi Inkubator Bersama Perguruan Tinggi Sebagai Bentuk Pengabdian Terhadap Masyarakat Dalam Perkembangan Iptek,” *ADI Pengabdi. Kpd. Masy.*, vol. 1, no. 2, pp. 23–31, 2021.
[19] T. Ayuninggati, E. P. Harahap, and R. Junior, “Supply Chain Management, Certificate Management at the Transportation Layer Security in Charge of Security,” *Blockchain Front. Technol.*, vol. 1, no. 01, pp. 1–12, 2021.
[20] B. Mardisentosa, U. Rahardja, K. Zelina, F. P. Oganda, and M. Hardini, “Sustainable Learning Micro-Credential using Blockchain for Student Achievement Records,” in *2021 Sixth International Conference on Informatics and Computing (ICIC)*, 2021, pp. 1–6.
[21] W. Setyowati, R. Widayanti, and D. Supriyanti, “Implementation Of E-Business Information System In Indonesia: Prospects And Challenges,” *Int. J. Cyber IT Serv. Manag.*, vol. 1, no. 2, pp. 180–188, 2021.
[22] T. Wahyuningsih, F. P. Oganda, and M. Anggraeni, “Design and Implementation of Digital Education Resources Blockchain-Based Authentication System,” *Blockchain Front. Technol.*, vol. 1, no. 01, pp. 74–86, 2021.
[23] D. Amany and A. Desire, “Pembelajaran Interaktif berbasis Gamifikasi guna Mendukung Program WFH pada saat Pandemic Covid-19,” *ADI Bisnis Digit. Interdisiplin J.*, vol. 1, no. 1, pp. 48–55, 2020.
[24] N. N. Halisa, “Peran Manajemen Sumber Daya Manusia" Sistem Rekrutmen, Seleksi, Kompetensi dan Pelatihan" Terhadap Keunggulan Kompetitif: Literature Review,” *ADI Bisnis Digit. Interdisiplin J.*, vol. 1, no. 2, pp. 14–22, 2020.
[25] P. O. A. Sunarya and N. Lutfiani, “Analisis Sistem Sertifikasi Profesi Untuk Pengembangan Kompetensi Mahasiswa,” *ADI Bisnis Digit. Interdisiplin J.*, vol. 1, no. 1, pp. 70–77, 2020.