Borneo International Journal eISSN 2636-9826; Vol. 7 (2); 2024; 27-40
Published by *Majmuah Enterprise*
www.majmuah.com

Study on Machine Learning Implementation in Cybersecurity for Security Defend and Attack

AZRAN ABDUL RAZAK
Jabatan Teknologi Maklumat dan Komunikasi (JTMK), Politeknik Tuanku Syed Sirajuddin (PTSS), *Arau, Perlis, Malaysia*

HELMY HANYFF HAIRUDIN RUZAILI
School of Computing, College of Arts and Sciences Universiti Utara Malaysia, Sintok, Kedah, Malaysia

MOHAMAD FADLI ZOLKIPLI
School of Computing, College of Arts and Sciences Universiti Utara Malaysia, Sintok, Kedah, Malaysia

Email: azranar@ptss.edu.my, mzulhelmin@utm.my

Received: April 03, 2024
Accepted: May 01, 2024
Online Published: June 04, 2024

Abstract

This comprehensive study explores the utilization of Machine Learning (ML) in the field of cybersecurity, emphasizing its substantial contribution to both defensive and offensive strategies. In contrast to conventional rule-based methodologies, machine learning systems can dynamically adjust to changing threats by acquiring patterns and anomalies from vast datasets. This study investigates the defensive utilization of machine learning (ML) in threat detection, anomaly identification, and security breach prediction. Additionally, it examines the offensive applications of ML, wherein attackers exploit vulnerabilities by applying advanced ML techniques. The study additionally examines the pragmatic implementations of machine learning (ML) in cybersecurity, specifically emphasizing a range of tools such as DeepExploit, Scikit-learn, Metasploit, Nmap, and antivirus software. An assessment is conducted to evaluate the defensive capabilities of Intrusion Detection Systems, firewalls, Security Information and Event Management systems, and email security solutions that utilize Machine Learning. Machine learning in these domains signifies a pivotal advancement in cybersecurity tactics, empowering firms to address cyber risks better.

Keywords: Machine Learning; Cybersecurity Attack; Cybersecurity Defends

1. Introduction

Machine Learning (ML) has emerged as a powerful tool in the realm of cybersecurity, offering advanced capabilities in both defence and offence strategies. Unlike traditional rule-based approaches, ML systems can dynamically adapt to evolving threats by learning patterns and anomalies from vast datasets. In cybersecurity, the focus on ML rather than Artificial Intelligence (AI) is deliberate. While AI is a broader concept encompassing machines that can perform tasks requiring human intelligence, ML specifically involves systems learning from data to make predictions or decisions. In the context of cybersecurity, ML's emphasis on pattern recognition, anomaly detection, and predictive analytics makes it particularly well-suited for handling the dynamic and complex nature of cyber threats. This study targeted to delve into the multifaceted application of Machine Learning in cybersecurity, exploring its role in both defensive and offensive strategies. On the defensive side, ML is employed for threat detection, anomaly identification, and the prediction of potential security breaches. Robust ML algorithms can analyze network traffic patterns, detect abnormal behaviours, and proactively defend against various cyber threats. Conversely, ML is increasingly being explored for offensive purposes, where attackers leverage sophisticated techniques to exploit vulnerabilities and breach security measures. Understanding the duality of ML in cybersecurity is crucial for developing comprehensive defence mechanisms and staying ahead of emerging cyber threats.

The organization of this paper is structured to provide a comprehensive exploration of the implementation of Machine Learning in Cybersecurity for security defence and attack purposes. The initial sections will delve into the fundamental principles of Machine Learning in the context of cybersecurity, distinguishing it from broader AI concepts. Following this, the study will explore the practical applications of ML in cybersecurity, focusing on its utilization for security defence mechanisms and its potential role in cyber-attacks. Each section will be accompanied by relevant case studies, examples, and insights derived from existing literature and real-world scenarios. By comprehensively addressing both defensive and offensive applications, this paper aims to contribute to the evolving discourse on the role of Machine Learning in shaping the future of cybersecurity strategies.

2. Literature review (background, history, issues/Threats

2.1 Background

2.1.1 Early Cybersecurity Challenges

The roots of cybersecurity challenges date back to the early days of computing, where concerns mainly revolved around unauthorized access and data manipulation. As technology progressed, so did the sophistication of threats, leading to the development of foundational defence mechanisms like firewalls and antivirus programs.

2.1.2 The Emergence of Advanced Persistent Threats (APTs)

The late 20th century witnessed a paradigm shift with the emergence of Advanced Persistent Threats (APTs). These prolonged and sophisticated attacks aimed at compromising sensitive information posed a significant challenge to traditional defence strategies. This era marked the beginning of a more targeted and persistent threat landscape. (Adel Asharani et. al,2019) say an APTs are typically carried out by financially supported attackers which are given the tools necessary to assault as long as the financial backing of the group requires. The attack ceases only upon detection or upon the financing organization receiving all necessary data.

2.1.3 Transition to Machine Learning

With the rise of Machine Learning (ML), the cybersecurity landscape experienced a transformative shift. ML's capability to adapt dynamically to evolving threats by learning patterns and anomalies from vast datasets addressed the limitations of rule-based approaches. The deliberate focus on ML, Unlike Artificial Intelligence (AI), cybersecurity represents a deliberate decision made to meet the unique requirements of the field. According to Arthur Samuel, machine learning is a field of research that enables computers to acquire knowledge and skills without relying on explicit programming. Arthur Samuel garnered much acclaim for his software designed for playing checkers. Machine learning is employed to enhance computers' ability to process data efficiently.

2.2 Current Cybersecurity Threat Landscape

2.2.1 Proliferation of Cyber Threats in the 21st Century

The 21st century has witnessed a significant surge in cyber threats, including a wide spectrum of activities such as state-sponsored hacking and ransomware attacks. The rapid growth of the Internet of Things (IoT) has led to heightened interconnectivity, expanding potential vulnerabilities and presenting novel challenges for defensive measures.Cyber threats have evolved to become more adaptive, leveraging sophisticated techniques that often bypass traditional security measures. According to (Muhammad Fakhrul Safitra, Muharman Lubis, and Hanif Fakhrurroja,2023) By continuously evolving and adapting to new threats, an evolutionary method of cybersecurity helps companies become more resilient against cyberattacks.

2.2.2 The Need for Machine Learning in Cybersecurity

In the current landscape, the cat-and-mouse game between defenders and attackers is more intricate than ever. Machine Learning's emphasis on pattern recognition, anomaly detection, and predictive analytics has become indispensable for addressing the dynamic and complex nature of contemporary cyber threats. This study aims to explore how ML applications can enhance both defensive and offensive strategies, providing a nuanced understanding of its pivotal role in fortifying cybersecurity measures. According to (Shaukat et.al, 2020) Because machine learning (ML) can learn from past events and promptly adapt to current attacks, it is one potential means of acting fast against such attacks.

3. Machine Learning implement on Cybersecurity

3.1 Overview Machine Learning in Cyber Attack and Defends

Both new challenges and inventive defend strategies have been introduced by the application of machine learning to cybersecurity. It is essential to include ML features to instruments utilized in both offensive and defensive cybersecurity methods to integrate machine learning into cyberattacks and defend. Because of this adaptation, machine

learning techniques can be used to improve defenses against new attacks and enhance the destructive potential of cyberattack weapons.

4. Cyberattack tools employ Machine Learning to attack

4.1 DeepExploit

An exploit's main goal is to use a particular vulnerability to carry out harmful actions. This can entail collecting sensitive data, running arbitrary code, or getting illegal access. According to (Kalle Kujanpää, Willie Victor and Alexander Ilin , 2024) A deep reinforcement learning agent called DeepExploit has been trained to automatically obtain first access through known vulnerabilities and exploits. On the Metasploit framework, it is constructed. After achieving successful penetration, it makes recursive efforts to reach more hosts within the local network of the input IP address. Instance after-exploitation, DeepExploit provides relatively little support as a security assessment framework for example, the application considers lateral movement to be an additional initial access activity.

4.2 Scikit-learn on ML Library

A Python machine-learning framework called Scikit-learn provides a variety of algorithms for cybersecurity jobs. Its adaptability is used by practitioners for dimensionality reduction, grouping, regression, and classification. This makes it possible to create customized models that handle certain security issues like anomaly detection and vulnerability severity prediction. Security experts may create tailored machine-learning solutions for improved cybersecurity defence because of the library's versatility and smooth connection with Python's larger ecosystem. According to (Fabian Pedregosa et. al ,2011) Scikit-learn exposes a wide range of supervised and unsupervised machine learning algorithms with a uniform, task-oriented interface, making it easy to evaluate approaches for a specific application. Since it requires the scientific Python ecosystem, it can be readily integrated into programs that go beyond the traditional purview of statistical data analysis.

4.3 Metasploit

Security experts, ethical hackers, and penetration testers may find, exploit, and evaluate vulnerabilities in computer systems with the help of Metasploit, an open-source penetration testing platform. Metasploit, created by Rapid7, offers an extensive toolkit for offensive security and testing for penetration. According to (Ovidiu Valea and Ciprian Oprisa, 2020) The complexity of this issue derives from the diversity of networks and systems, necessitating constant technique adaptation. The method is predicated on determining the features of the system, looking for vulnerabilities already there, and using machine learning to choose the best attack. The model offers exploits from the Metasploit toolkit and was trained using data gathered from compromised PCs on the "Hack theBox" learning platform. There are two types of automated and manual testing.

4.4 NMAP

Nmap, which stands for "Network Mapper," is an open-source program that is useful for security scanning and network discovery. Nmap is a popular tool used by network engineers, system administrators, and security specialists to evaluate and examine network hosts, services, and vulnerabilities. It was created by Gordon Lyon, better known by his stage name, Fyodor. According to (Arun S and Dr. Bijimol T K,2021) These tools are designed to help attackers locate potential attack routes and information about a target that they may exploit to compromise the device. This information gives the attacker potential contact details, which makes it very helpful in social engineering attempts.

4.5 Antivirus

The term "anti-virus," or "anti-virus," refers to a vital part of computer security that is used to safeguard computers against malware, or harmful software. Antivirus software is a first line of security for both consumers and enterprises as the digital world changes and more complex cyberthreats appear daily. Identifying, stopping, and getting rid of dangerous malware from computer systems is the main job of antivirus software. This includes spyware, trojan horse, worms and adware, among other harmful programs that could compromise the accessibility, integrity, and privacy of data. According to (Asamoah, 2021), antivirus software is a type of application designed to detect, stop, and remove malware infections on desktops and laptops, networks, and IT systems. Specialized security software called antivirus

software seeks to provide greater protection than what the underlying operating system (like Windows or Mac OS X) provides. It's utilized as a preventative measure most of the time.

5. Open Source Cyberattack tools employ Machine Learning to attack

The combination of cyberattack tools and artificial intelligence, particularly machine learning, has changed the way security risks are recognized and taken advantage of in the ever-changing field of cybersecurity. This investigation explores the use of machine learning in many well-known open-source cyberattack tools, emphasizing its potential applications and consequences for offensive security tactics. These technologies, which range from sophisticated network scanning utilities to automated penetration testing frameworks, highlight the increasing sophistication of cyber threats and the tactical adaptability of attackers in the digital sphere. According to (Calix et. al,2020).CyberSecTK, a toolkit for cyber security, is a basic Python module designed to preprocess and extract features from data related to cyber security. An increasing amount of data must be handled automatically as the digital world grows. Cyber security experts have seen chances in the past few years to process and analyze their data with the use of machine learning techniques.

5.1 DeepExploit

DeepExploit is at the forefront of the nascent field of machine learning and cyber exploitation. DeepExploit, created as a deep reinforcement learning agent, represents a revolution in hacking techniques. It is self-learning and can find and attack vulnerabilities on its own, having learned on the Metasploit framework. This allows it to get first access. Even though penetration testing is its primary emphasis, its recursive approaches demonstrate a novel approach by aiming to expand the reach within local networks following a successful breach. According to (Maeda and Mimura, 2021) DeepExploit executes the most effective exploit for the specified service after port-scanning the target server. This is usually when the cyberattack starts. On the other hand, the suggested method automates the after-exploitation, or the actions taken after DeepExploit has completed its run.

5.2 Scikit-learn (Machine Learning Library)

When it comes to offensive cybercrime methods, the Python-based machine-learning framework Scikit-learn becomes a double-edged weapon beyond its defence applications. With so many methods available, it provides practitioners with the freedom to design original models for anomaly detection and vulnerability severity prediction. Thanks to its adaptability and simplicity of integration with the Python environment, security experts may develop tailored machine-learning solutions that enhance the offensive capabilities of cyberattack weapons. According to (Raschka, Patterson, & Nolet, 2020) because it supports low-level libraries and clean high-level APIs, Python continues the most popular language for scientific programming, data analysis, and machine learning. This increases productivity and performance.

5.3 Metasploit

A cornerstone in the arsenal of security professionals, Metasploit, an open-source penetration testing platform, adopts a machine learning-driven approach. By leveraging data from compromised systems on the "Hack the Box" learning platform, Metasploit's model sifts through exploits within its toolkit, adapting techniques to the diverse networks and systems encountered. This integration of machine learning enables the platform to intelligently select the most effective attack strategies, emphasizing a dynamic and adaptive offensive methodology. According to (Quilantang et.al, 2021) because of its ease of use, the researcher claimed that Metasploit is among the greatest penetration testing tools out there. Furthermore, they said that there are still a lot of programs available for everyone to look over and select from.

5.4 Nmap

While traditionally recognized for its role in network scanning and discovery, Nmap, an open-source tool, inadvertently aids attackers in locating potential attack routes. Although not explicitly designed for offensive purposes, Nmap provides attackers with crucial insights into network topology and potential vulnerabilities. This information serves as a strategic foundation for threat actors, facilitating the identification of exploitable targets. According to, (Quilantang et.al, 2021) Nmap, often known as Network Mapper, is a potent open-source utility for network discovery and vulnerability assessment. It's widely considered one of the greatest free security tools out there. Nmap is a highly recommended tool for network discovery and vulnerability scanning by both security experts and ethical hackers due to its vital features and ease of use.

5.5 Antivirus

Even the defensive front of cybersecurity is not exempt from the influence of machine learning. Antivirus software, a cornerstone in safeguarding against malware, now integrates machine learning algorithms for enhanced threat detection. By analyzing patterns and behaviours, antivirus tools proactively defend against evolving cyber threats, exemplifying the bidirectional impact of machine learning on both offensive and defensive cyber strategies. According to (D. Santos,2021) When the majority of individuals research PC protection, some people ponder if buying antivirus software is essential to safeguarding their PC info and whether free software suffices.

5.6 Machine Learning in Cyber Defends

It is essential to protect machine learning technologies' cybersecurity from continually evolving threats. It is essential to implement strong encryption mechanisms and to perform regular updates and continual monitoring. Guarded cybersecurity protocols guarantee the confidentiality and integrity of sensitive data that these products process. The creation of intrusion detection systems (IDS) in (1990) is a prime example of the ways machine learning gets started in cybersecurity. Cyberbullying detection, anomaly and fraud detection, spam filtering, analysis of malware and detection, intruder detection, zero-day attack detection, Internet of Things assaults, threat analysis, and many other types of applications are all described by (Sarker, 2022).

5.7 Intrusion Detection System

As a first line of defend, intrusion detection systems (IDS) keep an eye on network activity and look for unusual or suspicious activity. Once the system finds patterns or behaviours that correspond with established algorithms that point to possible security breaches or risks, it can send out warnings or notifications. Administrators or security staff are then notified of the abnormalities found by these alerts. This increases the network's overall security posture by enabling a prompt reaction to detect, analyze, and mitigate any security threats. According to (Mouli et.al, 2023) raw data for sensors is provided by three primary information sources such as audit trails, syslog, and the IDS knowledge base. Syslog may contain information on system file settings, user permission, and other things. The foundation for future decisions is provided by this data. Modern IDS uses advanced machine learning techniques, such as ensemble methods and deep neural networks, to identify complex patterns that may be signs of upcoming cyberattacks. Because these systems are good at learning from large datasets, they can detect abnormalities in network activity and adapt to evolving attack techniques. Furthermore, quick identification and reaction to security issues are made possible by real-time analysis capabilities, which also operate as a preventative measure against new threats. According to (Naeem et.al,2022) The effectiveness of various machine learning strategies is required since it plays an important role in enhancing IDS performance. Classification algorithms play a crucial part in helping the IDSs differentiate between multiple forms of attacks. A dynamic and essential component of contemporary cybersecurity is using machine learning in intrusion detection systems (IDS), which provides a strong defense against the constantly changing landscape of cyberattacks.

5.8 Firewall

As the primary barrier in network security, firewalls meticulously monitor incoming and outgoing network traffic based on predefined rules. Acting as the initial line of defense, firewalls scrutinize data packets and make decisions to allow or block them, depending on established security protocols. In the event of detecting irregular patterns or potential threats through heuristic analysis or predefined algorithms, firewalls can generate alerts or notifications. These notifications are then relayed to administrators or security personnel, enabling swift responses to identified security risks. The vigilance of firewalls serves to fortify the network's security posture by promptly identifying, analyzing, and mitigating potential threats, thereby safeguarding the integrity of the network. According to (Abu Al-Haija & Ishtaiwi, 2021) to prevent cyber threats from entering the network, the filtration procedure is usually carried out by comparing the transmitted packets against specified instructions and rules. Thus, the incoming packet is either “allowed,” “denied,” or “dropped/reset” by the firewall system. In the realm of network security, modern firewalls leverage advanced machine learning techniques, including ensemble methods and deep neural networks, to discern intricate patterns indicative of potential cyber threats. Their proficiency in learning from extensive datasets enables the detection of anomalies in network traffic, facilitating adaptation to evolving attack methodologies. According to (Applebaum, Gaber, & Ahmed, 2021) the complication of defining rules makes it more challenging to adapt signature-based systems to counter new threats. Machine learning may be a key component in the solution to this issue. Real-time analysis capabilities empower these firewalls not only to swiftly identify and respond to security issues but also to proactively

thwart emerging threats. The implementation of machine learning techniques into modern firewall systems is becoming a more important and dynamic component of cybersecurity, providing a strong defense against the constantly changing cyberattack landscape.

5.9 Security Information and Event Management (SIEM)

Serving as a comprehensive security solution, Systems for Security Information and Event Management (SIEM) are essential to network defense. SIEM platforms meticulously aggregate and analyze security events and information from diverse sources within an organization's IT infrastructure. As the central hub for security monitoring, SIEM scrutinizes logs, correlates events, and applies predefined rules to detect potential security incidents. Acting as a sophisticated nerve center, SIEM systems not only identify irregular patterns or potential threats through heuristic analysis but also generate alerts or notifications. These notifications are then transmitted to administrators or security personnel, facilitating prompt responses to identified security risks. The vigilance of SIEM fortifies the overall security posture by swiftly identifying, analyzing, and responding to potential threats, contributing to the robust integrity of the network. According to (Alturkistani & El-Affendi, 2022) The fundamental method relies on machine learning supervised learning, wherein characteristics taken from web proxy network records are employed to identify malware integrated into HTTP and HTTPS (beaconing) traffic. SVM, random forest (RF), decision trees (DT), and logistic regression are some of the models that are employed. In the domain of cybersecurity, contemporary Security Information and Event Management (SIEM) systems harness cutting-edge machine learning techniques, incorporating ensemble methods and deep neural networks. These sophisticated methodologies empower SIEM to adeptly discern intricate patterns that may signify potential cyber threats within vast datasets of security events. According to (Adabi Raihan Muhammad, Parman Sukarno, & Aulia Arif Wardana, 2023)the system performs testing and can spot a denial-of-service attack. Once the DoS attack is identified, the system might issue a warning. The SIEM system's dashboard is used to show every attack detection alert. The researcher uses open-source SIEM to simulate the attack. Under the application of machine learning, SIEM improves its capacity to identify abnormalities in network behaviour, enabling quick responses to change attack tactics and approaches.In addition to improving threat detection, the integration of cutting-edge machine learning into SIEM presents it as a flexible and reliable solution for recognizing and addressing complicated cybersecurity issues.

5.10 Antivirus

Operating as a pivotal defend layer, various types of antivirus solutions diligently monitor system activities, scrutinizing for any indications of unusual or suspicious behaviour. Signature-based antivirus software identifies known malware by comparing file signatures with a database of recognized threats. Behaviour-based antivirus, on the other hand, analyses software behaviour for anomalous activities that might indicate malicious intent. Upon detecting patterns aligned with predefined algorithms suggestive of potential security threats, the antivirus system issues warnings or notifications. Security administrators are promptly alerted to these anomalies, enabling swift responses for detection, analysis, and mitigation of security risks, thereby enhancing the overall security posture of the system. According to (Bostani & Moonsamy, 2024) state that the results show the suggested adversarial approach is a flexible black-box attack that doesn't assume anything about the target detectors, such as the ML algorithms or the malware detection features. It can also function well in a variety of attack scenarios. Modern antivirus software incorporates advanced machine learning techniques, including ensemble methods and deep neural networks, to identify intricate patterns indicative of emerging cyber threats. Proficient in learning from extensive datasets, these systems excel at detecting abnormalities in system behaviour and adapting to evolving attack methodologies. Three unique machine learning (ML) algorithms, Decision Trees, Support Vector Machines, and N-gram have been carefully selected, according to (Nor Zakiah Gorment, 2023) the selection is grounded in the current performance outcomes of malware detection, which has demonstrated a remarkable 100% accuracy rate. The inclusion of these algorithms reflects a strategic approach to leverage their strengths and capabilities in enhancing the accuracy and efficacy of malware detection within the experimental framework. These findings collectively demonstrate the adaptability and strategic prowess of ML techniques in addressing cybersecurity challenges. The integration of machine learning represents a dynamic and indispensable facet of contemporary antivirus solutions, providing robust defense mechanisms in the ever-evolving landscape of cybersecurity challenges.

5.11 Email Security

Serving as the frontline defends, email security systems employ vigilant monitoring of electronic communications, scrutinizing for any anomalies or suspicious activities. Upon detecting patterns aligning with established algorithms

indicative of potential security risks, these systems promptly issue warnings or notifications. Security administrators are then alerted to these irregularities, facilitating swift responses for detection, analysis, and mitigation of security threats. This proactive approach significantly bolsters the overall security posture of the email system. According (Esra Altulaihan et.al, 2023) compared several machine learning algorithms and discovered that, in terms of accuracy, logistic regression performs better than stochastic gradient descent, naïve Bayes, support vector machines and random forest. Their trials on benchmark datasets demonstrate that, with an excellent accuracy rate of 91.9%, the logistic regression approach comes out as the most accurate method. Modern email security systems integrate advanced machine learning techniques, including ensemble methods and deep neural networks, to discern intricate patterns that might signify imminent cyber threats. Proficient in learning from extensive datasets, these systems excel at detecting abnormalities in email behaviour and dynamically adapting to evolving attack techniques. Regrettably, the study conducted by Md Yasin and Azmi (2023) reveals a limitation in the current machine learning algorithms employed for recognizing phishing emails. The researchers note that these techniques struggle to effectively identify ongoing phishing scams, emphasizing the need for extensive manual feature engineering to enhance their capabilities (Md Yasin & Azmi, 2023) While machine learning algorithms play a crucial role in detecting phishing emails, the dynamic landscape of phishing threats often requires the integration of manual solutions, such as intensive feature engineering, to stay ahead of emerging tactics and enhance the effectiveness of the overall security strategy. The combination of automated machine learning techniques and manual interventions is often essential for robust and adaptive defense against evolving phishing scams.

6. Paid Cyberattack tools employ Machine Learning to attack.

6.1 Antivirus

The importance of antivirus software in the constantly changing field of cybersecurity cannot be emphasized. As the first line of protection against the growing number of complex cyber threats, antivirus protection is a crucial component of computer security. In the current digital era, antivirus software emerges as the sentinel protecting both individuals and businesses since malicious software and viruses continuously test the security of computer systems. This talk explores the world of premium cybersecurity solutions, particularly those that use machine learning to strengthen our comprehension of how cutting-edge technologies are used in the fight against ever-changing cyber threats. According to (Thomas and Nachamai, 2017) Compared to free antivirus software, paid antivirus provides greater flexibility and comprehensive security features like parental controls and identity theft protection. Typically, expensive antivirus software consists of a few extra functions beyond suites, which are intended to be one-stop shops for protection. The largest problem with using free antivirus software is that there is typically no technical assistance available; customers must solve this on their own.

Table 1: Comparison table comparing Machine Learning (ML) integrated security products attack aspect.

Product	Operating System	Pricing	Type	User Administration	Latest Product Examples
Deep Exploit	Windows, Linux	Paid	Penetration Testing	Intermediate	CylancePROTECT ,Darktrace CrowdStrike Falcon Sophos, Intercept X, SentinelOne
Scikit-learn	Windows, Linux	Free and Paid	Machine Learning	Easy to Intermediate	TensorFlow,PyTorch ,Apache Spark MLlib ,H2O.ai,Dask-ML
Metasploit	Windows, Linux	Free	Penetration Testing	Easy to Expert	Rapid7 InsightVM, Cobalt Strike,Canvas, ,Nexpose,Armitage
Nmap	Windows, Mac, Linux	Free	Network Scanning	Easy to Intermediate	Zenmap,OpenVAS Wireshark

Antivirus	Windows, Mac, Linux	Free and Paid	Endpoint, Cloud	Easy to Intermediate	Bitdefender, Norton, Sophos, Kaspersky, Virus Total

Table 1 shows a detailed comparison of various cybersecurity products that integrate machine learning in their architecture to enhance their capabilities. The table is categorized into several attributes:

Operating System: Specifies compatible of the product within different operating systems.

Pricing: Indicates whether the product is available for free or if it involves a purchase.

Type: Describes the primary focus or category of the security product.

User Administration: Assesses the complexity of user administration, ranging from easy to expert.

Latest Product Examples: Represents some of the latest and well-known products in each category. For commercial antivirus software, examples could include Norton, McAfee, Bitdefender, etc.

7. Cyberattack tools employ Machine Learning to Defend aspects.

Below is a table comparing Machine Learning (ML) integrated security products defend (IDS, Firewall, SIEM, Antivirus, and Email Security) based on operating system compatibility, pricing, product type, user administration complexity and the latest popular product examples:

Table 2: Comparing Machine Learning Cybersecurity Tools Across Key Features

Product	Operating System	Pricing	Type	User Administration	Latest Product Examples
IDS	Windows, Linux	Free and Paid	Network-based, Host-based	Intermediate	Darktrace, Vectra AI
Firewall	Windows, Linux	Free and Paid	Network-based, Next-Gen	Intermediate	pfSense, Sophos XG Firewall, Fortinet FortiGate
SIEM	Windows, Linux	Varies	Software, Cloud	Expert	Splunk Enterprise, Elastic SIEM, IBM QRadar
Antivirus	Windows, Mac, Linux	Free and Paid	Endpoint, Cloud	Easy to Intermediate	Bitdefender, Norton, Sophos, Kaspersky
Email Security	Windows, Mac, Linux	Free and Paid	Cloud, On-Premises	Easy to Intermediate	Proofpoint, Mimecast, Barracuda Email Security Gateway

Table 2 shows a comprehensive overview of machine learning-integrated cybersecurity tools tailored to defensive operations. These products are detailed in terms of:

Operating System: Specifies compatibility of the product with different operating systems.

Pricing: Indicates whether the product is available for free or involves a purchase (varies based on features and licensing).

Type: Describes the primary focus or type of the security product.

User Administration: Assesses the complexity of user administration, ranging from easy to expert, based on the product's interface and functionalities.

Latest Product Examples: Represents some of the latest and well-known products in each category as of the latest available information.

8. Discussion Cyberattack tools employ Machine Learning to Attack aspect

8.1 DeepExploit

Using a deep reinforcement learning agent, DeepExploit is an effective tool for probing and examining network vulnerabilities. The fact that it works with other operating systems, such as Linux, macOS and Windows, indicates how versatile it is. Using well-known vulnerabilities, DeepExploit offers an automated method for obtaining first access to Windows, Mac, and Linux operating systems used by security experts. Because of its easy platform integration and build on the Metasploit framework, it is a great option for network security assessments and penetration testing.

8.2 Scikit-learn

A range of algorithms are provided by the Python machine-learning framework Scikit-learn to practitioners in the cybersecurity space. With compatibility with Windows, macOS, and Linux, this adaptable library is available to a wide range of users. Security professionals may use it for several operating systems activities as grouping, classification, and dimensionality reduction. The library's seamless interaction with the Python environment adds to its allure by making it possible to create machine-learning solutions that are specifically targeted to the cybersecurity requirements of users on various platforms.

8.3 Metasploit

One platform that stands out as an extensive toolbox for offensive security and penetration testing is Metasploit, an open-source penetration testing program. Because of its smooth operation on Windows, macOS, and Linux, this application offers security professionals, ethical hackers, and penetration testers accessibility and flexibility across a variety of operating systems. Metasploit accommodates a wide range of user preferences, be it a preference for the command-line capabilities of Linux, the dependability of Mac computers, or the Windows interface. Its feature-rich toolbox, ability to constantly adapt to new approaches, and use of machine learning for attack optimization make it an invaluable resource for cybersecurity experts across a range of platforms.

8.4 Nmap

The open-source tool Nmap, which stands for "Network Mapper," is frequently used for network discovery and security scanning. Because of its interoperability with Windows, macOS, and Linux, system administrators, network engineers, and security experts may use this program on a variety of operating systems. With its extended capability for Windows-based systems, Nmap offers users a powerful tool for assessing and investigating network hosts, services, and vulnerabilities. Nmap continues to be a well-liked option for evaluating possible attack paths and obtaining vital data for cybersecurity needs, regardless of operating system either Linux or macOS.

8.5 Antivirus

An essential part of computer security is antivirus software, which guards against malicious programs and viruses. Antivirus software, which is compatible with both Windows and macOS, is an essential first line of protection for personal computers, networks, and IT systems. Although antivirus programs are less common on Linux systems, they are nevertheless accessible for users who want an extra security measure. Users who are concerned about security on Windows and macOS may benefit from the detection, avoidance, and elimination of harmful malware, which protects the confidentiality, availability, and integrity of data in the quickly changing digital environment.

9. Discussion Cyberattack tools employ Machine Learning to Defend aspect

9.1 Intrusion Detection Systems (IDS)

Leverage machine learning to bolster their defend mechanisms against cyberattacks. By employing advanced algorithms, IDS can analyze network traffic patterns in real time, swiftly identifying and flagging suspicious activities indicative of potential security threats. Machine learning enhances the IDS's capability to adapt and evolve with emerging attack methodologies, enabling more accurate and proactive threat detection.

9.2 Firewalls

Integrate machine learning algorithms to fortify their defend against cyber threats. By analyzing vast amounts of network data, firewalls can discern complex patterns and behaviours associated with malicious activities, allowing them to make informed decisions regarding network traffic flow. Machine learning enables firewalls to dynamically adjust their filtering rules and policies in response to evolving threats, enhancing the overall security posture of the network.

9.3 Security Information and Event Management (SIEM)

Machine learning is used by systems to improve their capacity to identify security incidents and execute appropriate action. By analyzing large volumes of security event data, SIEM systems can identify anomalous behaviour patterns and prioritize critical security events for further investigation. In order to provide enhanced detection and mitigation techniques, machine learning algorithms enable SIEM systems to continually gain knowledge from and adapt to new threats.

9.4 Antivirus software

Integrates machine learning algorithms to enhance its capability to defend against malware and other cyber threats. By analyzing file characteristics and behaviour patterns, antivirus programs can detect and block malicious software in real time, protecting systems and data from potential harm. Machine learning enables antivirus software to improve its detection accuracy over time by learning from past encounters with malware, thereby staying ahead of emerging threats.

9.5 Email security

Solutions leverage machine learning algorithms to enhance their ability to detect and mitigate email-based cyber threats such as phishing attacks and malware distribution. By analyzing email content, sender behaviour, and other contextual factors, email security systems can identify suspicious emails and prevent them from reaching users' inboxes. Machine learning enables these systems to continuously refine their threat detection capabilities based on evolving email threat landscapes, ensuring effective protection against email-based cyber-attacks.

10. Discussion Comparison cybersecurity base machine learning tool on attack and defend

Table 3 shows the comparison of machine learning-based cybersecurity tools designed for both attacking and defending within the digital security realm. It lays out key criteria for each tool, highlighting their versatility and functionality in the context of security operations:

Table 3: Machine Learning Cybersecurity Tools for Attack and Defense Compared Across Key Criteria

Attack	Defend
Operating System: Cybersecurity solutions must be compatible with several operating systems like Windows, macOS, and Linux, making it versatile for many practitioners. Tools like Scikit-learn's machine learning capabilities on Windows, macOS, and Linux, Metasploit, a powerful penetration testing tool, work smoothly on Windows, macOS, and Linux, giving security experts and	**Operating System:** This criterion emphasises the compatibility of the security product with diverse operating systems. For example, if an antivirus program is compatible with Windows, Mac, and Linux, it implies that users on these operating systems can avail themselves of the security provided by the product. In the context of security appliances, it's essential to have a

ethical hackers freedom. Nmap's interoperability with Windows, macOS, and Linux enables cross-OS network scanning and vulnerability assessment, making it useful in cybersecurity studies. Antivirus software generally targets Windows and macOS, limiting its availability to these popular operating systems. This constraint shows the need for wider compatibility to manage cybersecurity risks in other situations.	strong and current operating system to guarantee dependability and effectiveness, especially when incorporating new machine learning-based algorithms. Regular updates and bug fixes are crucial to sustain the security and efficiency of these devices. To strengthen the resilience of the products against emerging cyber threats and vulnerabilities, security providers can prioritise operating system upgrades and bug patches, as well as ensure compatibility with different operating systems.
Pricing: Pricing models affect cybersecurity solutions' accessibility and functionality for consumers with varying budgets. DeepExploit uses a subscription model to offer penetration testing-specific functionality. Scikit-learn offers free and premium versions, ensuring machine learning capabilities for anyone. A notable free and open-source platform is Metasploit. It offers many penetration testing and offensive security technologies without cost. Nmap remains free and accessible, making it a useful network scanning tool for cybersecurity specialists and hobbyists looking to improve their online security. Antivirus software comes in free and paid versions, with paid subscriptions offering premium features and support. This meets user preferences for protecting digital assets from evolving cyber threats.	**Pricing:** This feature pertains to the availability of the security product, whether it is offered at no cost or requires a monetary transaction. Certain products may provide fundamental features at no cost, although more sophisticated functionalities and user-friendly management through a graphical user interface may necessitate a subscription or single purchase. The pricing of the product can differ depending on the features offered and the licensing type in place. The cost is crucial, necessitating meticulous preparation and allocation of funds to ensure that each type and aspect satisfy enhances the organization's cybersecurity stance. When choosing between free or premium security programs, it is crucial to carefully consider the features offered in relation to the fees involved, furthermore, it is imperative to comprehend the licensing type and ascertain any associated subscription costs in order to effectively plan the budget and allocate resources accordingly.
Type: Penetration testing and vulnerability exploiting tool DeepExploit is well-known, cybersecurity specialists can assess and improve system defenses with its resources. The extensible Scikit-learn is a machine learning framework specifically created to identify possible network vulnerabilities and plays a crucial role in creating personalised cybersecurity models. Metasploit is a commonly utilised software for conducting penetration testing, which utilises offensive security techniques to imitate cyber assaults and detect weaknesses within computer systems. Nmap is employed by security professionals to conduct scans and assessments of network vulnerabilities, hence enhancing the overall integrity and security of the system. Antivirus software is employed to identify and mitigate malware threats in order to protect computers. The incorporation of these technologies into cybersecurity processes serves to bolster the system's defensive capabilities. In order to enhance cybersecurity protocols, it is imperative to implement a complete approach that encompasses routine penetration testing, thorough machine learning analysis, wide network scanning, and continuous efforts to detect malware.	**Type:** This particular criterion serves to clarify the primary focus or categorization of the security product, hence assisting users in understanding the main purpose of the tool. The purpose of an Intrusion Detection System (IDS) is to conduct targeted surveillance. The machine learning framework Scikit-learn is essential for detecting possible dangers in a network, whereas antivirus software is designed to safeguard endpoints from malicious software. A prudent strategic decision would be to select a security system that incorporates machine learning algorithms alongside efficient tools. The process of integration facilitates the optimisation of security management, resulting in a decrease in the need for manual labour and resource allocation. The implementation of this complete approach not only serves to safeguard systems from a diverse array of cyber threats, but also contributes to the improvement of operational efficiency. Through the utilisation of machine learning algorithms, the system's ability to identify and address risks is greatly enhanced, enabling it to effectively identify and mitigate the impact of emerging and changing threats.
User Administration: For intermediate to expert cybersecurity experts and penetration testers, DeepExploit offers specialised user administration.	**User Administration:** This evaluates the intricacy of administering users within the programme, spanning from simple to advanced. An antivirus software with a

Beginning and experienced data scientists can use Scikit-learn's easy to somewhat advanced user management. Metasploit supports novice to advanced users. Designed for system administrators, network engineers, and security specialists, Nmap targets easy to sophisticated user administration. Antivirus software offers simple to moderately complicated administration features for normal users and IT managers. DeepExploit and antivirus software have varying user administration complexity. This makes them easy to utilise for cybersecurity specialists and regular users, improving cybersecurity resilience.	user-friendly interface is appropriate for those with rudimentary computer skills, whereas a SIEM system, specially designed for professionals, may require more intricate setups. The level of intricacy involved in user administration inside a security solution directly impacts the requisite technical proficiency for efficient management. Due to variations in organisational size and security needs, it may be necessary to augment the technical team in order to accommodate more advanced solutions. This guarantees the most efficient utilisation of resources and improves the effectiveness of cybersecurity.
Latest Product Examples: DeepExploit is proven effective in cybersecurity by CyberSecTK, Darktrace, CrowdStrike Falcon, and Sentinel One. Recent Scikit-learn products like TensorFlow, PyTorch, Apache Spark MLlib, H2O.ai, and Dask-ML demonstrate its importance in cybersecurity machine learning. Popular Metasploit tools like Rapid7 InsightVM, Cobalt Strike, Nexpose, and Armitage demonstrate its adaptability and efficiency as a penetration testing platform. Nmap's recent products, Zenmap, OpenVAS, and Wireshark, demonstrate its importance in network scanning and vulnerability evaluation. Finally, antivirus software like Bitdefender, Norton, Sophos, Kaspersky, and VirusTotal show how endpoint security solutions protect systems from malware. DeepExploit, Scikit-learn, Metasploit, Nmap, and antiviral software have a wide range of user administration difficulties, making them easy to use across technical skill levels. Overall cybersecurity resilience improves.	**Latest Product Examples:** This category showcases the latest and most renowned items in each security sector. Darktrace and Vectra Al are prominent instances of intrusion detection systems (IDS), exemplifying the most recent breakthroughs in intrusion detection technology based on the newest accessible information. Certain contemporary items may necessitate instruction and a significant amount of time to become proficient in, particularly when customising them to conform to the requirements of the organisation. Although new solutions frequently have sophisticated machine learning capabilities and need minimal human participation, effectively integrating them with pre-existing systems might provide difficulties. Hence, it is imperative to prioritise solutions that possess the potential to scale and extend, guaranteeing compatibility with forthcoming security requirements and simplifying seamless implementation and management procedures.

11. Conclusion

Machine learning empowers cyber attackers with advanced evasion, reconnaissance, and exploitation tactics, enabling targeted and undetected attacks. This poses significant challenges for cybersecurity defence as attackers leverage AI-driven techniques to automate tasks and adapt to defences. Conversely, the integration of machine learning in cybersecurity tools enhances threat detection, scalability, and efficiency. It enables proactive defence strategies, automates processes, and fosters continuous improvement, bolstering organizations' ability to effectively detect and respond to cyber threats.

Acknowledgement

The authors would like to thank all members of the School of Computing who participated in this study. This study was carried out as part of the Hacking and Penetration Testing Project. This work was supported by Universiti Utara Malaysia

Reference

Adel Alshamrani, Sowmya Myneni, Chowdhary, A., & Huang, D. (2019). A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities. *IEEE Communications Surveys and Tutorials*, *21*(2), 1851–1877. https://doi.org/10.1109/comst.2019.2891891.

Muhammad Fakhrul Safitra, Lubis, M., & Hanif Fakhrurroja. (2023). Counterattacking Cyber Threats: A Framework for the Future of Cybersecurity. *Sustainability (Basel)*, *15*(18), 13369–13369. https://doi.org/10.3390/su151813369.

Shaukat, K., Luo, S., Vijay Varadharajan, Hameed, I. A., & Xu, M. (2020). A Survey on Machine Learning Techniques for Cyber Security in the Last Decade. *IEEE Access*, *8*, 222310–222354. https://doi.org/10.1109/access.2020.3041951.

Kalle Kujanpää, Victor, W., & Ilin, A. (2021). Automating Privilege Escalation with Deep Reinforcement Learning. *ArXiv (Cornell University)*. https://doi.org/10.1145/3474369.3486877.

Pedregosa, F., Pedregosa@inria, F., Fr, Org, G., Michel, V., Fr, B., … Passos, A. (2011). Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot Edouard Duchesnay. *Journal of Machine Learning Research*, *12*, 2825–2830. Retrieved from https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https:/.

Ovidiu Valea, & Ciprian Oprisa. (2020, September 3). Towards Pentesting Automation Using the Metasploit Framework. Retrieved April 3, 2024, from ResearchGate website: https://www.researchgate.net/publication/347188530_Towards_Pentesting_Automation_Using_the_Metasploit_Framework.

Arun, S., & Bijimol, T. K. (2021). A research work on information gathering tools. In *Proceedings of the National Conference on Emerging Computer Applications (SNCECA-2021)* (p. 118). Amal Jyothi College of Engineering. https://doi.org/10.5281/zenodo.5101265.

Asamoah, H. (2019). Antivirus software versus malware. Retrieved April 3, 2024, from Архів кваліфікаційних робіт (2019-2020 ÷ 2022-2023 н.р.) website: https://jarch.donnu.edu.ua/article/view/10531.

Calix, R. A., Singh, S. B., Chen, T., Zhang, D., & Tu, M. (2020). Cyber Security Tool Kit (CyberSecTK): A Python Library for Machine Learning and Cyber Security. *Information*, *11*(2), 100–100. https://doi.org/10.3390/info11020100.

Maeda, R., & Mimura, M. (2021). Automating post-exploitation with deep reinforcement learning. *Computers & Security*, *100*, 102108–102108. https://doi.org/10.1016/j.cose.2020.102108.

Raschka, S., Patterson, J., & Nolet, C. (2020). Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information*, *11*(4), 193–193. https://doi.org/10.3390/info11040193.

Quilantang, K. A. G., Rivera, J. A. C., Pinili, M. V. M., Magpantay, A. J. N. R., Blancaflor, E. B., & Pastrana, J. R. A. M. (2021). Exploiting Windows 7 vulnerabilities using penetration testing tools. In The 2021 9th International Conference on Computer and Communications Management. https://doi.org/10.1145/3479162.3479181.

Santos, D. (2021, November). Comparison of Paid Subscription vs Freeware Software on Antivirus Program. Retrieved April 3, 2024, from Hawaii.edu website: https://dspace.lib.hawaii.edu/items/2a4b1eb5-6307-4b24-8d4a-e2952db09e98.

Thomas, R., & M. Nachamai. (2017). Performance Investigation of Antivirus – A Comparative Analysis. *Oriental Journal of Computer Science and Technology*, *10*(1), 201–206. Retrieved from https://www.computerscijournal.org/vol10no1/performance-investigation-of-antivirus-a-comparative-analysis.

Sarker, I. H. (2022). Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. *Annals of Data Science (Print)*, *10*(6), 1473–1498. https://doi.org/10.1007/s40745-022-00444-2.

K. Chandra Mouli, B. Indupriya, D. Ushasree, Ch.V. Raghavendran, Rawat, B., & Bhukya Madhu. (2023). Network Intrusion Detection using ML Techniques for Sustainable Information System. *E3S Web of Conferences*, *430*, 01064–01064. https://doi.org/10.1051/e3sconf/202343001064.

Naeem, S., None Aqib Ali, None Sania Anam, & Ahmed. (2022). Machine Learning for Intrusion Detection in Cyber Security: Applications, Challenges, and Recommendations. *Innovative Computing Review*, *2*(2). https://doi.org/10.32350/icr.0202.03.

Abu Al-Haijaa, Q., & Ishtaiwia, A. (2021). Machine learning based model to identify firewall decisions to improve cyber-defense. International Journal of Advanced Computer Science and Applications, 11(4). https://doi.org/10.18517/ijaseit.11.4.14608.

Applebaum, S., Gaber, T., & Ahmed, A. (2021). Signature-based and Machine-Learning-based Web Application Firewalls: A Short Survey. *Procedia Computer Science*, *189*, 359–367. https://doi.org/10.1016/j.procs.2021.05.105.

Hilala Alturkistani, & El-Affendi, M. A. (2022). Optimizing cybersecurity incident response decisions using deep reinforcement learning. *International Journal of Power Electronics and Drive Systems (Online)*, *12*(6), 6768–6768. https://doi.org/10.11591/ijece.v12i6.pp6768-6776

Adabi Raihan Muhammad, Parman Sukarno, & Aulia Arif Wardana. (2023). Integrated Security Information and Event Management (SIEM) with Intrusion Detection System (IDS) for Live Analysis based on Machine Learning. *Procedia Computer Science*, *217*, 1406–1415. https://doi.org/10.1016/j.procs.2022.12.339.

Hamid Bostani, & Veelasha Moonsamy. (2023). EvadeDroid: A Practical Evasion Attack on Machine Learning for Black-box Android Malware Detection. *Computers & Security*, 103676–103676. https://doi.org/10.1016/j.cose.2023.103676.

Nor Zakiah Gorment, Selamat, A., Lim Kok Cheng, & Ondrej Krejcar. (2023). Machine Learning Algorithm for Malware Detection: Taxonomy, Current Challenges, and Future Directions. *IEEE Access*, *11*, 141045–141089. https://doi.org/10.1109/access.2023.3256979

Esra Altulaihan, Abrar Alismail, Rahman, H.sA. (2023). Email Security Issues, Tools, and Techniques Used in Investigation. *Sustainability*, *15*(13), 10612–10612. https://doi.org/10.3390/su151310612.

Yasin, S., & Hadi Azmi, I. (2023). *EMAIL SPAM FILTERING TECHNIQUE: CHALLENGES AND SOLUTIONS*. *101*(13). Retrieved from https://www.jatit.org/volumes/Vol101No13/6Vol101No13.pdf.