Intrusion Detection System Using Machine Learning

Swati[1], Dr. Manikamma Malipatil[2], Dr. Jyoti Neginal[3]

[1]M.Tech Student, Dept.of Computer Science and Engineering, Sharnbasva Univercity, Kalaburagi, Karnataka, India

[2]Associate Professor, Dept.of Computer Science and Engineering, Sharnbasva Univercity, Kalaburgi, Karnataka, India

[3]Associate Professor, Dept.of Computer Science and Engineering, Sharnbasva Univercity, Kalaburgi, Karnataka, India

Abstract— — **In order to safeguard our contemporary digital ecosystems, this project will look at the application of AI to the task of identifying cyber threats. For anomaly-based malware detection and network intrusion detection, it is crucial to evaluate ML-based classifiers and ensembles and figure out how to use these models in a network and mobile environment. We used the KDDCup-'99' dataset to train and evaluate a particular model that was the focus of our study. The model contained various strategies. Here, we demonstrate how to train a machine learning-based intrusion detection system (IDS) to mistake legitimate network traffic samples for hostile ones. In order to improve performance when confronted with new attacks, this data may also be used for training purposes. Using a signature-based intrusion detection system is also.**

***Keywords*—Network, security, AI, Cyber, Itrusion, treat**

I. Introduction

When malicious or suspicious behavior is detected on a network, an Intrusion Detection System (IDS) will notify the appropriate parties. Intrusion Detection Systems (IDSs) mostly identify and report anomalies, but some of them may also react to the finding of malicious activity or unusual traffic. Everything you need to know about an Intrusion Detection System is covered in this article.

Where does the term "intrusion detection system" come from?

If an intrusion detection system (IDS) detects anything out of the ordinary in a user's network traffic, it will alert them promptly. Network and system security software can identify malicious actions and stop them before they happen. It is standard practice to either alert management or centrally document any illegal activity or breaches utilising security information and event management system. IDS monitor a network or system for malicious actions and block access to the network or system from any unauthorized user, even if they are already within the system..

Methods Used by Intrusion Detection Systems (IDS)

- An IDS keeps tabs on a network's traffic in order to spot any questionable behavior.
- It searches for patterns and indications of unusual activity in the data streaming over the network.
- IDS searches for suspicious network activity by comparing it to database of known rules and patterns.
- Notify system administrator if intrusion detection system identifies something that corresponds to one of these patterns or rules.

- After receiving an alert, system administrator may analyze the situation and take measures to stop any more intrusion or harm.

Intrusion Detection System (IDS) Categorization

There are five distinct kinds of IDS:

- NIDS scan all device traffic for indicators of intrusion after they are installed at a specific spot within the network. All subnet communication is monitored and compared to a database of known threats in order to detect known assaults. When assaults or unusual activities are detected, administrators may get notifications. It is possible to identify intrusion attempts by installing network intrusion detection systems (NIDS) on the subnet, which includes firewalls.
- Host intrusion detection systems (HIDS): HIDS function on individual devices or hosts inside a network. A hardware intrusion detection system (HIDS) alerts the system administrator to any malicious or suspicious activity by monitoring the device's incoming and outgoing packets. By capturing a picture and comparing it to the previous snapshot, it verifies the present status of the system files. If the administrator suspects tampering with or deletion of the analytical system files, they will get a notice to investigate. Critical devices that are essential to the operation of the system and will not be altered in design are the ones that employ HIDS.

A system that detects intrusions

- In its permanent home at the server's front end, the Protocol-based Intrusion Detection System (PIDS) oversees and interprets the protocols that users and devices use to communicate with the server. As a safety measure, the web server adopts the associated HTTP protocol and routinely examines the HTTPS protocol stream. Given that HTTPS does not use encryption, this system must reside on this interface prior to directly contacting the web presentation layer.An APIDS is an agent or system that often resides inside a cluster of servers. For the purpose of detecting intruders, it intercepts and decodes data sent via application-specific protocols. Web server database transactions might benefit, one of which is monitoring the SQL protocol transmission to the middleware.
- When several intrusion detection methods are combined, the outcome is a hybrid system. By integrating information from both the host agent or system and the network itself, hybrid intrusion detection systems provide comprehensive picture of the network system. Overwhelmingly, the hybrid intrusion detection system outperforms its
- competitors. Hybrid IDS is exemplified by Prelude.

Ways to Avoid Intrusion Detection Systems

- One method is fragmentation, which is breaking the packet into smaller pieces. Since a malware signature is not possible in this case, it is hard to detect an infiltration.
- • Encoding packets using hexadecimal or Base64 might obfuscate harmful material from signature-based intrusion detection systems.
- To conceal an assault and evade detection, one might use traffic obfuscation, which makes the message more difficult to understand.
- Data integrity, secrecy, and privacy are just a few of the security benefits offered by encryption. Malware authors unfortunately use security features to obfuscate their assaults and evade detection.
- Reasons to Use IDS
- Identifies illegal activity: Intrusion detection systems may identify any questionable actions and notify the system administrator prior to any substantial harm being caused.
- • Boosts network performance: Intrusion detection systems may pinpoint any slowdowns in the network's performance, allowing for their remediation.

- By keeping tabs on network activity and producing reports, IDS may assist in satisfying compliance obligations.
- Insights into network traffic are generated by IDS, which may be used to uncover vulnerabilities and enhance network security.

Detection Method of IDS

- • Intrusion Detection Systems Based on Signatures: These systems identify intrusions by analyzing unique patterns in network traffic, which may be described by metrics like byte count or binary count. In addition,
- it identifies malware based on its known risky series of instructions. The IDS has found patterns, and such patterns are called signatures. When it comes to freshly found malware attacks, signature-based intrusion detection systems (IDS) have a hard time identifying them since their pattern is unknown. On the other hand, signature-based IDS are fantastic at discovering assaults whose pattern is already existing in the system.
- Method Based on Abnormalities: The rapid evolution of malicious software necessitated the introduction of anomaly-based intrusion detection systems to spot previously unseen attacks. It's core is a reliable activity model built using machine learning. We then evaluate all incoming data to this model and mark anything that doesn't fit as suspicious. Due to the fact that machine learning models can be customized to unique use cases and hardware settings, they beat signature-based intrusion detection systems when it comes to generalizability.

Intrusion detection and prevention systems mainly use one of three approaches: hybrid, signature-based, or anomaly-based. The IDS uses various methods to analyze data in search of potential intrusions. Using Anomalies to Detect Intrusions: The primary objective of intrusion detection systems that seek for anomalies is to identify instances when a system or network is behaving in an unusual manner.

II. Related Work

Regarding reference 15, see Shaukat et al. Artificial intelligence has several potential uses in the cybersecurity industry, including protecting networks, computers, mobile devices, and more. Because of their design and implementation, computer systems and networks are susceptible to cyberattacks and other forms of malicious cyber activity. Problems with setup, insufficient processes, and incompetent personnel are some of the ways in which computer network systems may be compromised. Attacks and threats, both internal and external to a network, are made more probable by these vulnerabilities. People in all sorts of professions are becoming more and more dependent on cyber technology. Any outside force capable of altering the functioning and behavior of a computer system or network via the deployment of a specific penetration technique is considered a threat. [16].

Preventing unauthorized access to computer systems, networks, and data is the primary goal of cybersecurity. [17].

The advent of viruses in 1970 marked the beginning of a long and contentious conflict between cybercriminals and security agencies [18].

With each passing second, these cyber security threats become more difficult to defend against. To address these security threats, academics are increasingly concentrating on the urgent need for innovative automated security solutions. Using automated machine learning algorithms is one of the most effective and efficient ways to find new cyberthreats. [19].

Machine learning techniques have been the subject of many scholarly articles. In their study, the authors compared support vector machines (SVMs) with the C 4.5 approach. [20].

We put the two algorithms through their paces in the face of four different kinds of network intrusions. According to the results, C 4.5 is better than SVM. The J48 tree, a version of the C 4.5 approach developed by Weka, yielded findings that corroborate this discovery.Among several other types of attacks, the writers of [21] researched a plethora of classification algorithms. While searching for the best algorithms, the authors took into account every conceivable kind of assault. The findings show that the leading algorithms for R2L attacks, U2R attacks, and PROBE assaults are Naive Bayes, Bayes Net, and One-R, in that order. Against this particular sort of denial-of-service assault, the majority of algorithms reportedly performed quite well. Staff experience in these specialized areas of cybercrime detection may be lacking; machine learning approaches might assist bridge this knowledge gap. tools. Furthermore, proactive tactics are necessary for detecting and countering the next wave of attacks, whether automatic or evolutionary. Machine learning (ML) is one approach to quickly responding to these types of assaults since it can learn from previous attacks and quickly counter future ones. Almost every part of cyber defense is now using some kind of machine learning or deep learning model to detect and prevent intrusions [22].

III. METHDOLOGY

Dataset Used

KDD Cup 1999

Presented at KDD-99, 5th International Conference on Knowledge Discovery and Data Mining, this dataset was utilized into this conference. Network intrusion detectors, or prediction models that can differentiate between "good" (i.e., normal) connections and "bad" (i.e., intrusions or attacks), were the focus of the competition. A spectrum of incursions mimicking a military network environment is the typical set of data to be checked out from this database.

SVM –

Support Vector Machines

Model for Support Vector Machines In the field of information security, Support Vector Machine (SVM) is often considered the gold standard when it comes to machine learning (ML) technologies for IDS. SVM data classification is based on the hyperplane's margin notation to split the data into two groups.

Decision Tree

One kind of supervised machine learning is the decision tree (DT), which uses recursive trees. A DT consists of 3 parts: a root or intermediate node, a route, and a leaf node.

Naïve Bayes

Naive Bayes (NB) classifiers deconstruct the conditional probability of the input data by using Bayes' theorem, also called Bayes' rule. problem under consideration. Nevertheless, this independence need is not satisfied by many attack kinds when it comes to cyber security.

1) Evaluation metrics

Machine learning makes use of a number of performance metrics—Accuracy, False Positive Rate, Precision, Recall—to assess efficacy of ML-based intrusion detection systems. When it comes to binary classification issues, nevertheless, Accuracy (AC) score is by far the most used statistic for evaluating model performance. Its definition is given by the following equation:

$$AC=(TP+TN)(TP+TN+FP+FN)$$

True Negative outcomes is represented by TN and the number of True Positive findings is represented by TP. Term "True Positive rate" describes the percentage of assaults that were correctly detected. True Negative will keep track of the number of correctly detected cases for each typical case. On the other hand, False Positive (FP) refers to the quantity of legitimate traffic that is mistakenly identified as an assault. The term "False Negative" (FN) describes the frequency with which incidents of attacks are wrongly seen as normal. It is common practice in intrusion detection to aim on FN reduction rather than FP reduction since FN is a more significant danger. When dealing with data from imbalanced classes, the Accuracy measure takes into consideration four additional measures. The F1 score, Precision, Recall, and False Positive Rate (FPR) are these metrics. The objective is to find positive instances of incursions as efficiently as possible.

1.*False Positive Rate* *(FPR)* The false positive rate (FPR) may be calculated by dividing the total number of false positives (FP) by the total number of real negatives. Give me another way of putting it:

$$FalsePositiveRate(FPR)=FP(FP+FN)$$

2.*Precision* Precision is the metric that determines how well a model performs by counting the number of times the model gets an instance's prediction right. One possible mathematical representation is the following equation:

$$Precision=TP(TP+FP)$$

3.*Recall/ Detection Rate* *(DR)* It is a metric for how well the ML model identifies True Positive cases. Recall also evaluates the model's precision in locating important data points. This is the rationale for its other names, such as True Positive Rate or Sensitivity. Below is a representation of the mathematical expression:

$$Recall/DR=TP(TP+FN)$$

4.*F1 Score* The total accuracy of the ML model is measured by the tradeoff between Recall and Precision, which takes both FP and FN into consideration. The following equation may be used to express the F1 score:

$$F1score=(2*Recall*Precision)(Recall+Precision)$$

-

IV. System Architecture

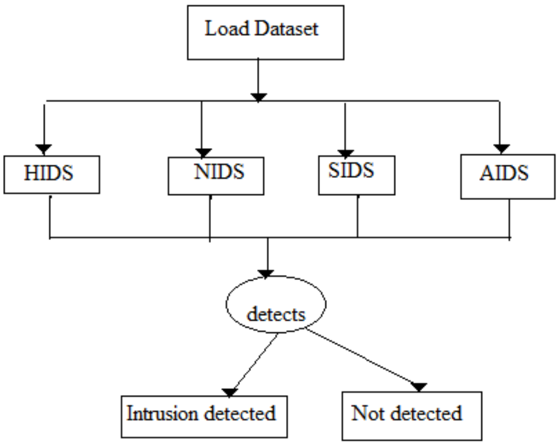


Fig-1: System Architecture

System divided into four categories: HIDS, NIDS, AIDS, SIDS to detect intrusion.

V. Implementation

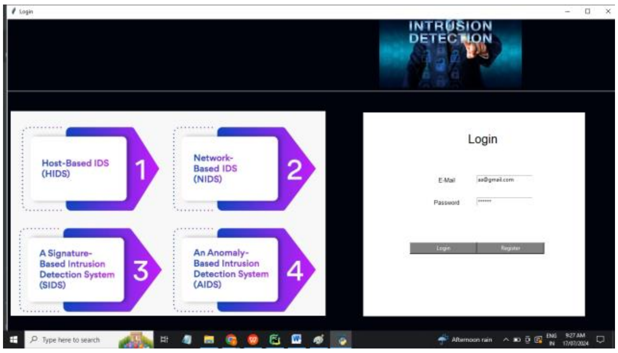


Figure 2 : Main Screen

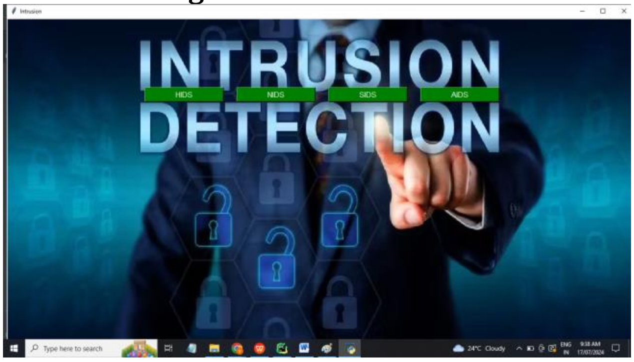
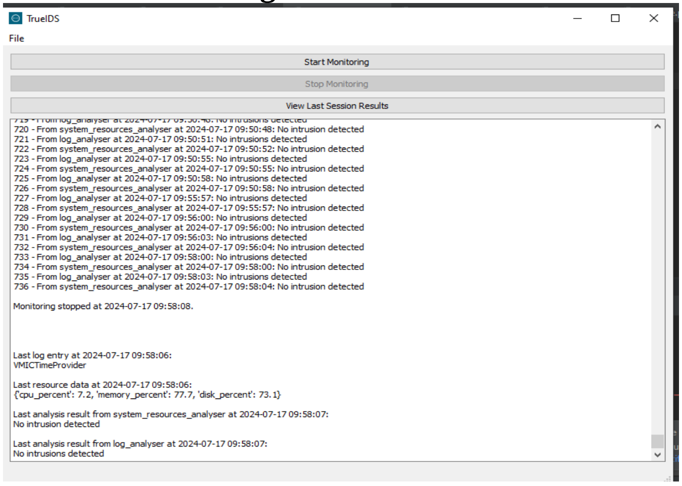
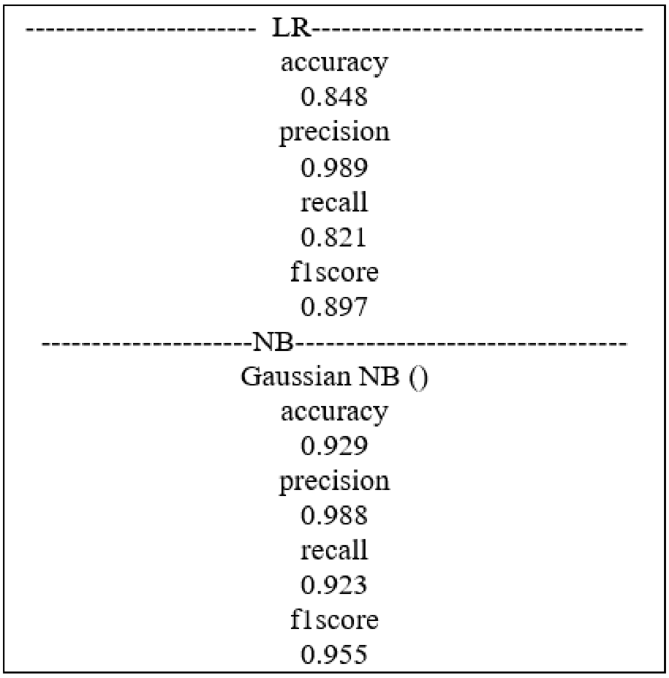


Figure 3: Menu



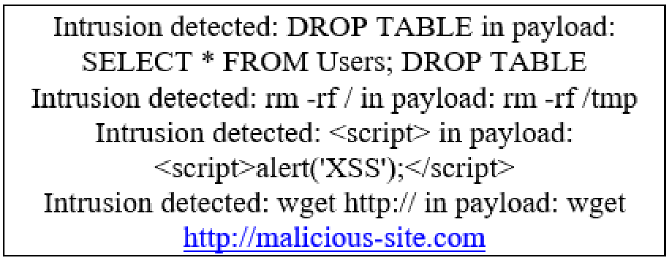
```
719 - From log_analyser at 2024-07-17 09:50:48: No intrusions detected
720 - From system_resources_analyser at 2024-07-17 09:50:48: No intrusion detected
721 - From log_analyser at 2024-07-17 09:50:51: No intrusions detected
722 - From system_resources_analyser at 2024-07-17 09:50:52: No intrusion detected
723 - From log_analyser at 2024-07-17 09:50:55: No intrusions detected
724 - From system_resources_analyser at 2024-07-17 09:50:55: No intrusion detected
725 - From log_analyser at 2024-07-17 09:50:58: No intrusions detected
726 - From system_resources_analyser at 2024-07-17 09:50:58: No intrusion detected
727 - From log_analyser at 2024-07-17 09:55:57: No intrusions detected
728 - From system_resources_analyser at 2024-07-17 09:55:57: No intrusion detected
729 - From log_analyser at 2024-07-17 09:56:00: No intrusions detected
730 - From system_resources_analyser at 2024-07-17 09:56:00: No intrusion detected
731 - From log_analyser at 2024-07-17 09:56:03: No intrusions detected
732 - From system_resources_analyser at 2024-07-17 09:56:04: No intrusion detected
733 - From log_analyser at 2024-07-17 09:58:00: No intrusions detected
734 - From system_resources_analyser at 2024-07-17 09:58:00: No intrusion detected
735 - From log_analyser at 2024-07-17 09:58:03: No intrusions detected
736 - From system_resources_analyser at 2024-07-17 09:58:04: No intrusion detected

Monitoring stopped at 2024-07-17 09:58:08.



Last log entry at 2024-07-17 09:58:06:
VMICTimeProvider

Last resource data at 2024-07-17 09:58:06:
{'cpu_percent': 7.2, 'memory_percent': 77.7, 'disk_percent': 73.1}

Last analysis result from system_resources_analyser at 2024-07-17 09:58:07:
No intrusion detected

Last analysis result from log_analyser at 2024-07-17 09:58:07:
No intrusions detected
```

Figure 4: HIDS



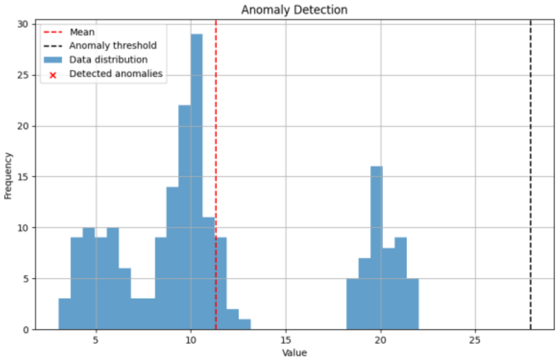
```
---------------------- LR--------------------------------
accuracy
0.848
precision
0.989
recall
0.821
f1score
0.897
--------------------NB-------------------------------
Gaussian NB ()
accuracy
0.929
precision
0.988
recall
0.923
f1score
0.955
```

Figure 5: NIDS



```
Intrusion detected: DROP TABLE in payload:
SELECT * FROM Users; DROP TABLE
Intrusion detected: rm -rf / in payload: rm -rf /tmp
Intrusion detected: <script> in payload:
<script>alert('XSS');</script>
Intrusion detected: wget http:// in payload: wget
http://malicious-site.com
```

Figure 6: SIDS



Graph 7: Anamoly detection

VI. Conclusion

New challenges for cyber security have emerged as a result of the tremendous advancements in information technology. New, more dependable, scalable, and flexible solutions are required due to the computational complexity of cyber assaults. Concern about cyber security has increased on a worldwide basis as a means to better detect and react to attacks. Due to their inability to detect novel and polymorphic attacks, the conventional security solutions of yesteryear are inadequate. Cybersecurity systems rely heavily on machine learning methods for a range of purposes. Machine learning and cyber security have been the subject of an explosion of publications over the last decade, thanks to the growing interest in these topics in academia, industry, and government. The focus of this study is on addressing cyber security concerns using an AI-based strategy. Current research indicates that the primary objectives of using AI in cyber security include detecting intrusions into networks, analyzing and classifying malware, and combating phishing and spam emails. Over time, DL became the standard practice there. Combining ML/DL with other smart approaches, such bioinspired methods, was also of interest to the researchers. These synergies keep the demand for more research alive while producing very promising results. Using machine learning for cyber security concerns was not their primary goal. Evasion could quickly mislead the ML model with malicious inputs. Reliable machine learning places more emphasis on the safe use of internet-based machine learning methods than on the model's accuracy or speed. In addition to outlining some of the primary challenges of applying machine learning approaches to cyber security, this presentation includes a thorough literature review. All of the above should be considered in future research.

REFERENCES

(Cavelty, Myriam Dunn, "The Routledge Handbook ofNew Security Studies," 154-162, 2018). To achieve effective and safe data capture for cloud-supported Internet of Things in smart grid, [2] Guan ZT, Li J, Wu LF, et al. The 2017 edition of the IEEE Internet of Things Journal looked at pages 1934–1944.

The authors of the paper "Big data analysis-based secure cluster management for optimized control plane in software-defined networks" (Wu et al., 2013) are responsible for the third element. ServManag, IEEE Transactions on, 15(1), 27–38.

Four, Buczak and Guven. An analysis of cyber security data mining and ML techniques "intrusion detection" (2015, 18, 1153–1176) in the IEEE Communications Surveys Tutorial.

García-Nieto, P.J.; Comesaña, C.I.; Torres, J.M. Cybersecurity using machine learning methods. Int. J.Mach. Discover. Cybern. 2019, 1–14.

[6] Shang, T.; Liu, J.; Guan, Z.; Bian, L. A survey on the intersection between machine learning and security concerns. Publication: Shenyang, China, 24-27 August 2018; pages: 158-165. This was part of the 2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR).

[7] Yang Xin, Kong Liu, Chen Yin, Li Yin, Zhu Huo, Gao Ming, Hou Huo, and Wang C. Cybersecurity using machine learning and deep learning techniques. The article is published in the 2018 edition of IEEE Access, volume 6, pages 35365–35381.

A study of deep learning algorithms for cyber security [8] Berman, D.S., Buczak, A.L., Chavis, J.S., and Corbett, C.L.News 2019, 10, 122.

Wickramasinghe, C.S., Marino, D.L., Amarasinghe, K., and Manic, M. [9] produced this work. Cyber-Physical System Security using Deep Learning: A Comprehensive Review. This is an excerpt from the published proceedings of the 2018 IEEE Industrial Electronics Society Annual Conference (IECON) held in Washington, DC, USA, from October 21st to the 23rd, 2018, pages 745 to 751.

Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., and Marchetti, M. (2010) published a study. Regarding the efficacy of artificial intelligence and deep learning in protecting against cyber threats. The 10th International Conference on Cyber Conflict (CyCon) took place from May 30th to June 1st, 2018 in Tallinn, Estonia, and the proceedings include pages 371-390.

Cybersecurity and AI: A comprehensive review (Li, J.H., 2011). Technical Advances in Electron Engineering 2018, 19, 1462–1474.

[12] Ivan Zelinka, Thanh Cong Truong, and Quoc Bao Diep. Symposium 2020, "Artificial Intelligence in the Cyber Domain: Offense and Defense," 12, 410.

[13] "A Survey of Artificial Intelligence in Cybersecurity," ACM SIGCHI Conference on Computational Science and Intelligence (CSCI) 2020, pp. 109-115, K. Morovat and B. Panda.

The article "A Study of Cyber Security using Machine Learning Techniques" was published in the International Journal of Innovative Technology and Exploring Engineering (IJITEE) in May 2019. The authors are R. Devakunchari, Sourabh, and Prakhar Malik. The ISSN number is 2278-3075.

*"A survey on machine learning techniques for cyber security in the last decade" (Shakat K et al., 2015). Volume 8, pages 222310-222354, 2020, published by IEEE Access.

In "Managing vulnerabilities of information systems to security incidents," published in 2003 in the Proceedings of the 5th international conference on Electronic commerce (pp. 348–354), F. Farahmand, S. B. Navathe, P. H. Enslow, and G. P. Sharp were the authors.

["Defining cybersecurity," by D. Craigen, N. Diakun-Thibault, and R. Purse, published in 2014 in the Technology Innovation Management Review, volume 4, number 10.

Citation: P. Szor [18] Pearson Education published it in 2005. The Science and Art of Defending Against Computer Viruses.

[19] "Analysis of machine learning techniques used in behavior-based malware detection," in 2010: IEEE Conference on Advances in Computing, Control, and Telecommunication Technologies, edited by I. Firdausi, A. Erwin, and A. S. Nugroho, pages 201-203, published in 2010.

[20] in Intruder Detection System Utilizing Data Mining Techniques with the C 4.5 Algorithm, Solanki and Dhamdhere (2015) published in the International Journal of Application or Innovation in Engineering and Management (IJAIEM), volume 4, issue 5, 2015.

[21] In the proceedings of the 11th Asia-Pacific Symposium on Network Operations and Management: Challenges for Next Generation Network Operations and Service Management (APNOMS '08), "Application of Data Mining to Network Intrusion Detection: Classifier Selection Model" was published (pp. 399-408). The authors of the paper are Nguyen and Choi. Beijing, China, in 808.

[22] In "Cyber Security: The Lifeline of Information and Communication Technology" (Springer, 2020), R. Prasad and V. Rohokale discuss the role of AI and ML in protecting networks and data.