Machine Learning Classifier Algorithms for Ransomware Lockbit Prediction

Ibrahiem M. M. El Emary[1,*], Khalil A. Yaghi[2]

[1,2] *Information Science Department - King Abdulaziz University, Jeddah, Saudi Arabia*

(Received: November 13, 2023; Revised: November 30, 2023; Accepted: December 15, 2023; Available online: January 29, 2024)

Abstract

Advanced virus known as ransomware has been spreading quickly in recent years, resulting in considerable financial losses for a variety of victims, including businesses, hospitals, and people. Modern host-based detection techniques need to first infect the host in order to spot abnormalities and find the malware. When the system is infected, it can already be too late because some of the assets have been exfiltrated or encrypted by the malware. On the other hand, as most ransomware families attempt to connect to command-and-control servers before to executing their damaging payloads, network-based methods can be helpful in detecting ransomware attacks. Therefore, one of the most important methods for early identification can be a detailed examination of ransomware network activity. This study presents a thorough behavioral analysis of the ransomware LockBit. In early 2022, ransomware, particularly targeting data on endpoints in Indonesia, was enough to horrify the news online. LockBit ransomware is one of the ransomwares that is particularly worrisome in Indonesia, so study is required to combat the ransomware. Static and dynamic analyses are used to study the ransomware; the former involves deciphering the portable executable (PE) file, while the latter involves actually running the ransomware. These analyses will reveal the impurity and resolve of the LockBit ransomware. Examine the running operations, the resources utilized, the network activities the ransomware performed, and the effect it had on the impacted operating system to try to build a scenario for preventative measures. The real effects of the ransomware-as-a-service (Raas) attacks conducted by the LockBit ransomware are demonstrated in this research. In this work, we describe an attribute selection-based system for identifying and avoiding ransomware that uses a variety of machine learning techniques, such as neural network-based frameworks, to classify the malware's security grade. We used a range of machine learning approaches, such as Decision Tree-DT, Random Forest-RF, Naive Bayes-NB, and Logical Regression-LR based classifiers, on a selected set of attributes for ransomware detection. The results of the study demonstrate that the Random-Forest predictor outperformed different classifiers by achieving the best accuracy, precision, recall, and F1-Score.

Keywords: Ransomeware LockBit, LockBit attacks, Decision Tree, Naïve Bayes, Logistic Regrssion as well as Random Forest

1. Introduction

Computer security is crucial in the current era of digitization since every piece of data contains crucial information. Attack structures have evolved in the current era of cybersecurity attacks from the previous attacking directly at the intended target, such as a company's server. By using malicious programs, innocent users can now become the object of an attack. The ways in which safety on computers is attacked as well as threatened had likewise evolved dramatically since the past [1]. Malware is commonly used to facilitate cybercrimes for a number of reasons, including locking or deactivating the device, stealing, erasing, or encrypting data, taking control of the device to attack additional companies or organizations, and receiving login information that grant access to the business's or organization's systems or services. Malware is also frequently used to attack other computers or devices. that you use and make use of these offerings for, but among all of these uses, there is software referred to as ransomware [2] that regularly encrypts your data and requests payment to decrypt it [2]. Additionally, ransomware is frequently used as a cover for Trojans or other types of malwares that function as backdoors so that victims are only directed to encrypted files without worrying about anything else, including other actions taken by the ransomware [3]. Ransomware frequently has the ability to enter machines on a network, do network discovery, and spread very quickly before immediately encrypting the computer.

Malicious software or assaults, such as those from the ransomware and malware families, persist in being an important danger to protection and has the ability to gravely harm networks, servers, websites, and smartphone apps across many companies and industries [4][6]. Most ransomware employs an unbreakable encryption technique that can only be

***Corresponding author: Ibrahiem M. M. El Emary (omary57@hotmail.com)**

DOI: https://doi.org/10.47738/jads.v5i1.161

deciphered by the attacker themselves in order to block and stop targeted victims from accessing computer data. In order to avoid irreparable damages caused by ransomware removal, victims are compelled to comply with the attacker's demands [7]. Data loss will result from refusal to comply with the attacker's request or failure to do so. The use of developing ransomware families, which are more challenging to remove after a ransomware infection, by attackers is made possible by modern technologies [8].

People all across the world are vulnerable to the sophisticated and diversified danger known as ransomware, which prohibits them from reaching their computers or data until a ransom payment is received [2]. The user's files are encrypted or the system screen is locked. According to attack tactics, the two primary varieties of ransomware are locker ransomware, which restricts availability of the processer or equipment, and crypto-ransomware, which prevents access to files or data [9]. After these attacks, it is quite difficult to turn around while paying the extortion. Event-based, statistical, and data-centric methodologies used in conventional ransomware detection are unsuccessful against it. Therefore, the scientific community should prioritize developing the highest level of ideal security and protection through the use of modern technology to guard versus such cunning hostile attacks.

2. Ransomeware Lockbit

LockBit is a type of ransomware attack in an ongoing series of cyberextortion attacks. Although it was once known as "ABCD" ransomware, it has since evolved into a serious threat in the setting of such extortion tools. LockBit is a sort of ransomware referred described as a "crypto virus" since it based its ransom demands on payments in cash in exchange for decryption. It typically concentrates on businesses and political institutions rather than people.

RaaS version Lockbit first surfaced in September 2019 under the name ABCD ransomware (because of its.abcd file extension). Accenture was infected by LockBit in July 2021, which stole internal data and encrypted systems that were later recovered from backups.

Ransomeware A specific piece of malware is known as LockBit, and the criminal organization that created it also goes by that name. In a business model known as ransomware as a service (Raas), the LockBit group also offers to sell its malware to other operators in exchange for money. The malware has been marketed as "the fastest encryption software in the world" on anonymous forums.

2.1. Various Stages of LockBit Attacks

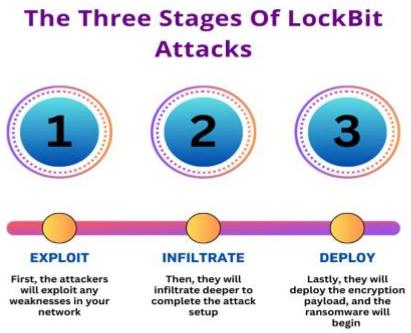

The Three Stages Of LockBit Attacks
1
2
3
EXPLOIT
First, the attackers will exploit any weaknesses in your network
INFILTRATE
Then, they will infiltrate deeper to complete the attack setup
DEPLOY
Lastly, they will deploy the encryption payload, and the ransomware will begin

Figure 1. Severalt LockBit Attacks Levels

There are essentially 3 phases to LockBit attacks:

phase 1: Exploit

phase 2: Infiltrate

phase 3: Deploy

phase 1: Take advantage of a network's flaws. The initial breach resembles other malicious attempts quite a bitPhishing is a type of social engineering technique where attackers ask for access credentials by posing as legitimate employees or authorities. Another method is to launch assaults using brute force on an organization's intranet services and network systems. Attack probes might be finished in a matter of days if the network is not configured properly.

When LockBit enters a network, the ransomware gets everything set up to spread its encrypting payload to as many devices as it can. An attacker might need to take a few extra precautions before they can make their last move, though.

phase 2: If more cover is required, move deeper to finish the offensive setup. From this point on, every activity is independently managed by the LockBit application. It is programmed to escalate privileges in order to obtain an attack-ready level of access by using so-called "post-exploitation" tools. To check for target feasibility, it also roots via access that is already open via lateral movement.

LockBit will now take any necessary precautions before launching the ransomware's encryption feature. In order to prevent system recovery, security software and other infrastructure must be disabled.

The purpose of intrusion is to prevent or slow down unassisted recovery to the point where paying the attacker's ransom is the only viable option. The victim is going to pay the ransom when they are desperate to resume normal business activities.

phase 3: Put the encryption payload into action. LockBit will start spreading to any computer it can touch once the network has been fully set up for its full mobilization. LockBit doesn't require much to finish this step, as was previously mentioned. To download and install LockBit, a only system unit with high access can send commands to other network units.

The encryption component will "lock" each system file. Only a unique key generated by LockBit's own decryption program can be used by victims to unlock their systems. Additionally, copies of a plain text ransom note file are left behind in each system folder by the operation. Instructions are given to the victim on how to restore their system, and some LockBit versions have even incorporated frightening blackmail.

The victim chooses what to do after each stage is finished. They might choose to pay the ransom and get in touch with LockBit's support team. It is not suggested to comply with their requests, nevertheless. There is no assurance for victims that the assailants will uphold their half of the contract.

2.2. LockBit Risk Categories

LockBit, the most recent ransomware epidemic, is extremely dangerous. We are unable to rule out the possibility that it will extend to a number of businesses and sectors, especially given the current boom in remote employment. Knowing the versions of LockBit will help you understand what you're up against.

Variant 1-. Abcd(extension)

In the original release of LockBit, files with the ".abcd" suffix are renamed. Additionally, each folder has a "Restore-My-Files.txt" file which carries a ransom note with demands and instructions for alleged recoveries.

Variant 2-. LockBit(extension)

This ransomware's present name was given to it when the second version that is now known to exist started using the ".LockBit" file extension. Despite some small backend modifications, victims will find that other aspects of this version appear to be largely unchanged.

Variant 3 -. LockBit version 2

The downloading of the Tor browser is no longer mentioned in the ransom specifications for the upcoming LockBit version. Instead, it sends them to another website using a regular internet connection.

Static analysis and dynamic analysis are the two categories into which the approaches used in this study to evaluate infection evaluation, persistence, and methods of avoiding LockBit ransomware may be divided.

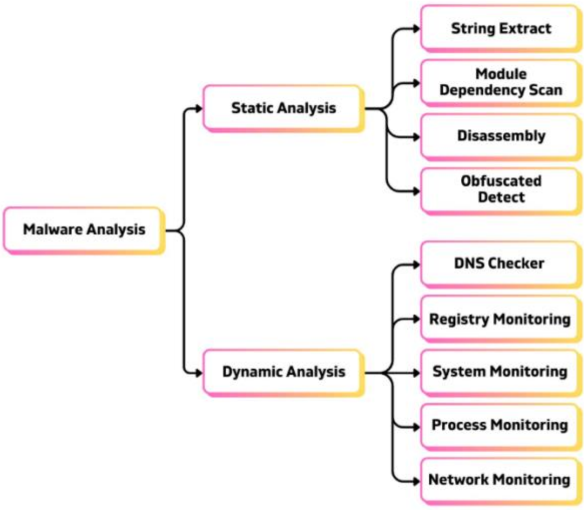


Figure 2. Analysis of Malware

2.2.1. Static Analysis

Static analysis is the first methodology employed, and it is used as the first step in identifying a file that may be malware so that it can be determined through analysis that it contains malicious code and falls under the category of malware. Several actions are taken at this point, including the following [10]:

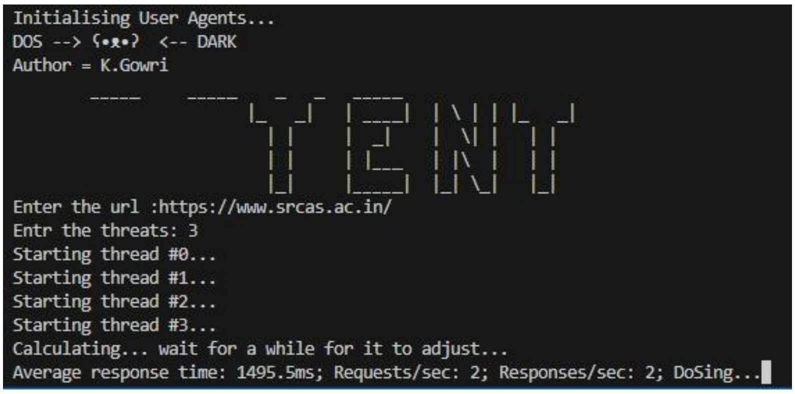
1. Static evaluation involves breaking down the target file, unpacking it, and comparing it to see if it fits into the malicious or harmless category.

2. The type of file to be analyzed is decided using hexadecimal analysis. The executable file can be classified as a Windows or DOS executable with a Part Executable (PE) file and has a hexadecimal code or a 4D 5A signature. Tools like IDAPro Free and PEid [11] are required to aid in the study of the file.

3. CRC32 is just one of the hash functions developed to detect data corruption during the procedure of transmission or storage, and analysis using the static in approach seeks to identify a file [12]. A file's CRC32 identification indicates whether it has undergone information transmission from its source to its intended destination.

2.2.2. Dynamic Analysis

Another method called dynamic evaluation, examines the infiltrated computer system connectivity together with its infrastructure, operating system, and computational tasks that affected machines carry out in order to determine the behavior aa well as output of the ransomware or infection. To investigate this, you'll need software such Process Hacker, Process Monitor, the software Wires and others.[13]



```
Initialising User Agents...
DOS --> ʕ•ᴥ•ʔ  <-- DARK
Author = K.Gowri
       _____     _____ _   _  _____
                      |_   _| | ____| | \ | | |_   _|
                        | |   | _|    |  \| |   | |
                        | |   | |___  | |\  |   | |
                        |_|   |_____| |_| \_|   |_|
Enter the url :https://www.srcas.ac.in/
Entr the threats: 3
Starting thread #0...
Starting thread #1...
Starting thread #2...
Starting thread #3...
Calculating... wait for a while for it to adjust...
Average response time: 1495.5ms; Requests/sec: 2; Responses/sec: 2; DoSing...
```

Figure 3. System screen when Thread Finding

3. Methodology

We used cutting-edge technology, like machine learning, to combat ransomware. The development of creative ransomware solutions can greatly benefit from the use of LockBit detection, a recent study area [7]. By implementing Machine Learning (ML) approaches, security is improved and malware, including ransomware, can be automatically detected through their dynamic behaviour [8]. For the categorization and detection of ransomware, algorithms Decision Tree-DT,Random Forest-RF, Naive Bayes-NB, and Logistic Regression -LR may be effective [9]. In this study, we undertake a thorough analysis and look into machine learning methods used for categorizing ransomware. The following are the paper's main contributions:

- We do a thorough analysis into categorizing ransomware and offer a framework by choosing a num of parameters for the procedure for developing a framework and utilizing NN-based architectures and conventional ML classifiers.

- Using reliable trials, we show how well the models work generally and contrast it with other approaches.

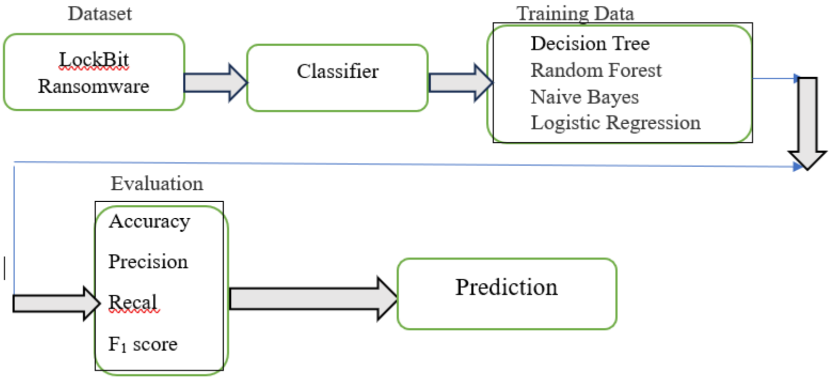


Figure 4. Ransomware LockBit Prediction using Machine learning models

3.1. Evaluation Metrics

Accuracy: The percentage of predictions that the algorithm accurately foretold is called efficiency. The following is the legal definition of accuracy:

$$Accuracy\ =\left(\frac{No\ Of\ Correct\ Prediction}{Totall\ no\ of\ Prediction}\right) \quad (1)$$

Additionally, accuracy can be determined using positives and negatives in the following order:

$$Accuracy=\frac{TP+FN}{TP+TN+FP+FN} \quad (2)$$

Recall: The proportion of all the positive samples with accurate positive predictions. Mathematically:

$$\text{Recall}=\frac{TP}{TP+FN} \quad (3)$$

Precision: a measurement of the proportion of positives out of all the positives expected that were actually successfully detected. Mathematically:

$$Precision=\frac{TP}{TP+FP} \quad (4)$$

Where TP stands for True Positive (amount of correctly predicted positive-outcomes) as well as FN for False Negative (number of incorrectly predicted positive-outcomes).

$F1\ score$: the harmonious combination of recall and precision. The accuracy measure for unbalanced data is inferior to the F1 score as a performance indicator.

$$\text{F1}=2X\frac{Precision\ X\ Recall}{Precision+Recall} \quad (5)$$

The biased harmonic mean of precision and memory is known as the F-Beta score, with a value of 1 denoting the best result (perfect precision and recall) and a value of 0 denoting the worst.

$$\text{F}\beta = (1+\beta^2)\,\frac{Precision\ X\ Recall}{\beta^2\ X\ Precision + Recall} \qquad (6)$$

When = 1, F-beta equals the F1 score. The precision and recall weights are set by the parameter. If we want to provide greater weight to precision, we can choose 1, whereas > 1 values give more weight to recall.

4. Result and Discussion

In this study, we used the classifiers DT, RF, NB, and LR to distinguish between genuine and ransomware LockBit samples. The models' performance is shown in Table 3 in terms of “accuracy, F1 score, recall, and precision”. The Random-Forest classifier surpasses extra models by scoring the maximum in terms of precision, F1score, recall, and accuracy out of 10. Regarding all of other performance indicators, it performs poorly. When compared to other classifiers, both the DT and NB classifiers perform reasonably well. However, LR falls short of other approaches in terms of F1score and recall values, yet the accuracy is respectable when compared with DT classifiers.

Table 1. Analysis of experimental data using different classifiers

	Accuracy	**Precision**	**Recall**	**F1 score**
Decision Tree (%)	9.8	9.4	9.4	9.8
Random Forest (%)	99.3	99.5	99.4	99
Naïve Bayes (%)	45	97	96	56
Logistic Regression (%)	96	89.6	89.5	94

4.1. Decision Tree Vs Random Forest:

For managing categorical and continuous data, a decision tree is utilized. It is a straightforward and useful decision-making diagram. It is far less likely to be impacted by outliers when using the random forest technique to generate numerous distinct decision trees and then average these forecasts [14][15][16]. RF Results are exact and accurate as a result.

Fig 5 Generally speaking, Random Forest outperforms Decision tree classifier in terms of performance. Although it has a lower generalization error rate than others, it is more noise-resistant and has higher performance characteristics.

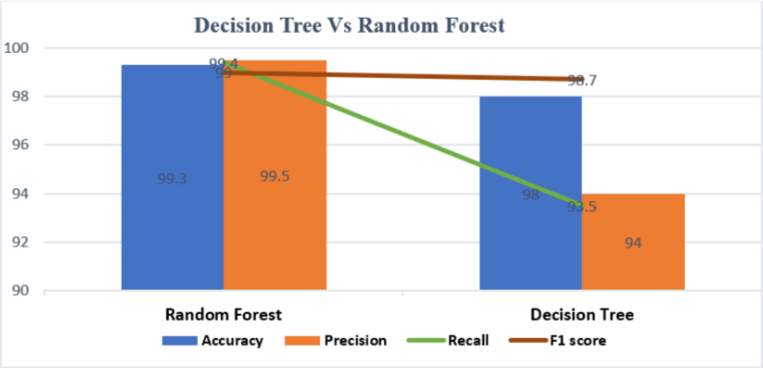


Figure 5. Performance of the Decision Tree and Random Forest models for prediction.

4.2. Naïve Bayes Vs Random Forest

To forecast labels or classify labels for data, there are many different classification techniques available. Random Forest and Nave Bayes are two of them. With the use of indices of accuracy, precision, F1 Score, and recall, both algorithms are capable of providing detailed descriptions of predictions [17][18][19].

Fig 6: According to our research, the Random Forest algorithm predicts with 99.3% accuracy, 99.5% precision, 99.4% recall, and 99% F1-Score. The aforementioned technique outperforms the Naive Bayes algorithm, which has an F1-Score of 56% and accuracy, precision, and recall values of 45%, 97%, and 96% respectively.

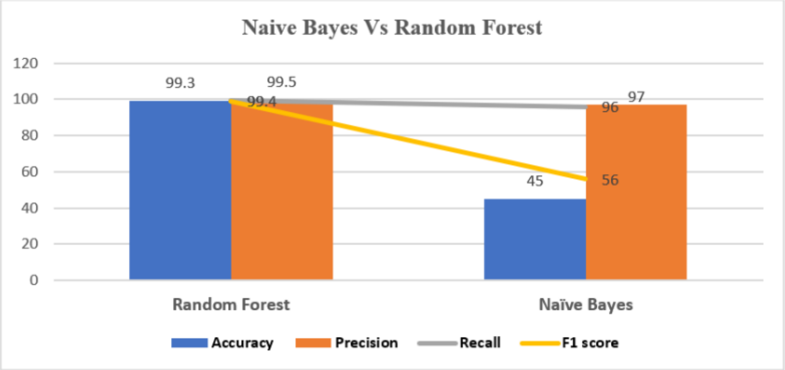


Figure 6. Performance of the Naïve Bayes and Random Forest models for prediction.

4.3. Logistic Regression Vs Random Forest:

A logistic function is used in logistic regression to make predictions [20][21]. Compared to Random Forest, it is less accurate and resilient, but computationally faster [22][23]. Fig. 7: According to our research, the Random Forest algorithm predicts with 99.3% accuracy, 99.5% precision, 99.4% recall, and 99% F1 Score. The aforementioned approach outperforms the log regression algorithm, which achieves 96% accuracy, 89.6% precision, 89.5% recall, and 94% F1 Score.

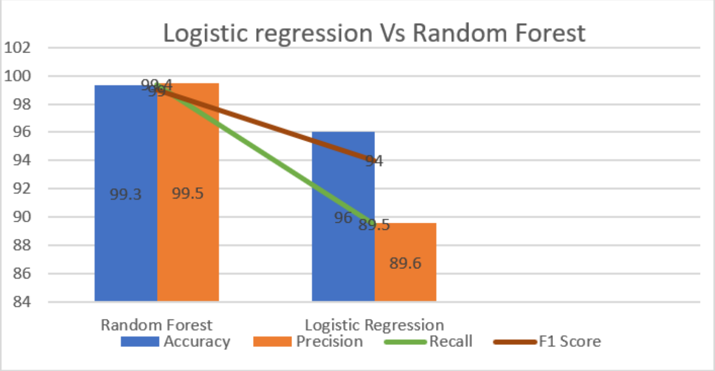


Figure 7. Performance of the Logistic Regression and Random Forest models for prediction.

5. Conclusion

Financial institutions, companies, and people are all facing major security threats from malware, including ransomware. A self-driving system must be built to precisely classify as well as notice and detect ransomware and reduce the threat of harmful actions. In this study, we employed a variety of machine learning methods, particularly neural network-based classification algorithms, and developed an inventive framework based on feature selection for the successful classification and detection of ransomware. On a ransomware dataset, we used the framework along with all of the trials, and we assessed the models' performance using a thorough comparison of DT, RFA, NB (Naïve Bayes), Logistic Regression classifiers. The experimental evaluation of the suggested detection method shows that it is very good at tracking ransomware network activity and has a low false positive rate, valid extracted characteristics, and high detection accuracy.

6. Declarations

6.1. Author Contributions

Conceptualization: I.M.M.E.; Methodology: K.A.Y.; Software: I.M.M.E.; Validation: I.M.M.E. and K.A.Y.; Formal Analysis: I.M.M.E. and K.A.Y.; Investigation: I.M.M.E.; Resources: K.A.Y.; Data Curation: K.A.Y.; Writing Original Draft Preparation: K.A.Y. and I.M.M.E.; Writing Review and Editing: K.A.Y. and I.M.M.E.; Visualization: I.M.M.E.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. J. Haber and B. Hibbert, "Ransomware," in *Privileged Attack Vectors*, vol. 1, no. 1, pp. 1-13, 2018.

[2] NCSC, "Mitigating malware and ransomware attacks," *Natl. Cyber Secur. Cent*., vol. 1, no. 1, pp. 1-12, 2020.

[3] H. Alshaikh, N. Ramadan, and H. Ahmed, "Ransomware Prevention and Mitigation Techniques," *Int. J. Comput. Appl.*, vol. 177, no. 40, pp. 1-8, 2020, doi: 10.5120/ijca2020919899.

[4] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A Survey on Machine Learning Techniques for Cyber Security in the Last Decade," *IEEE Access*, vol. 8, no. 1, pp. 222310–222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[5] N. Shah and M. Farik, "Ransomware-Threats, Vulnerabilities and Recommendations," *Int. J. Sci. Technol. Res.*, vol. 1, no. 1, pp. 1-8, 2017.

[6] M. J. Hossain Faruk et al., "Malware Detection and Prevention using Artificial Intelligence Techniques," in *Proc. - 2021 IEEE Int. Conf. Big Data, Big Data 2021*, vol. 1, no. 1, pp. 1-12, 2021.

[7] F. Noorbehbahani, F. Rasouli, and M. Saberi, "Analysis of machine learning techniques for ransomware detection," in *Proc. 16th Int. ISC Conf. Inf. Secur. Cryptology, Isc. 2019*, vol. 1, no. 1, pp. 128–133, 2019, doi: 10.1109/ISCISC48546.2019.8985139.

[8] U. Adamu and I. Awan, "Ransomware prediction using supervised learning algorithms," in *Proc. - 2019 Int. Conf. Futur. Internet Things Cloud, FiCloud 2019*, vol. 1, no. 1, pp. 57–63, 2019, doi: 10.1109/FiCloud.2019.00016.

[9] K. Savage, P. Coogan, and H. Lau, "The Evolution of Ransomware," *Res. Manag.*, vol. 54, no. 5, pp. 59–63, 2015.

[10] S. Gadhiya, K. Bhavsar, and P. D. Student, "Techniques for Malware Analysis," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 4, pp. 1-12, 2013.

[11] N. M. Hai, M. Ogawa, and Q. T. Tho, “Packer identification based on metadata signature,” *Proceedings of the 7th Software Security, Protection, and Reverse Engineering / Software Security and Protection Workshop*, vol. 1, no. 1, pp. 1–13, 2017. doi:10.1145/3151137.3160687.

[12] S. YusirwanS, Y. Prayudi, and I. Riadi, "Implementation of Malware Analysis using Static and Dynamic Analysis Method," *Int. J. Comput. Appl.*, vol. 117, no. 6, pp. 17-23, 2015, doi: 10.5120/20557-2943.

[13] L. Xu, D. Zhang, N. Jayasena, and J. Cavazos, "HADM: Hybrid Analysis for Detection of Malware," in *Lecture Notes in Networks and Systems*, vol. 16, no. 1, pp. 1-13, 2018.

[14] S. Raja and K. Venkatesh, "Using Honey Pot Technique Ransomeware Get Detected," *2023 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, vol. 1, no. 1, pp. 1-4, 2023, doi: 10.1109/ICCCI56745.2023.10128365.

[15] A. Ajiono and T. Hariguna, “Comparison of Three Time Series Forecasting Methods on Linear Regression, Exponential Smoothing and Weighted Moving Average,” *Int. J. Informatics Inf. Syst.*, vol. 6, no. 2, pp. 89–102, Mar. 2023

[16] P. K, B. Nataraj and P. Duraisamy, "An Investigation on Attacks in Application Layer Protocols and Ransomeware Threats in Internet of Things," *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, vol. 1, no. 1, pp. 668-672, 2023, doi: 10.1109/ICACCS57279.2023.10112669.

[17] C. C. Moreira, D. C. Moreira, and C. de Sales Jr., “Improving ransomware detection based on portable executable header using xception Convolutional Neural Network,” *Computers & Security*, vol. 130, no. 1, pp. 103265–103273, 2023. doi:10.1016/j.cose.2023.103265

[18] I. Nordat, B. Tola, and M. Yasin, “The Effect of Work Motivation and Perception of College Support on Organizational Commitment and Organizational Citizenship Behavior in BKPSDM, Tangerang District”, *Int. J. Appl. Inf. Manag.*, vol. 2, no. 3, pp. 40–49, Feb. 2022.

[19] A. Huertas Celdrán et al., “Behavioral fingerprinting to detect ransomware in resource-constrained devices,” *Computers & Security*, vol. 135, no. 1, pp. 103510–103518, 2023. doi:10.1016/j.cose.2023.103510

[20] A. Maharini Adiandari, “Financial Performance Innovation Since Digital Technology Entered Indonesian MSMEs”, *Int. J. Appl. Inf. Manag.*, vol. 2, no. 1, pp. 50–58, Dec. 2021.

[21] E. B. Karbab, M. Debbabi, and A. Derhab, “SwiftR: Cross-platform ransomware fingerprinting using hierarchical neural networks on hybrid features,” *Expert Systems with Applications,* vol. 225, no. 1, pp. 120017–120024, 2023. doi:10.1016/j.eswa.2023.120017

[22] F. Ali Faraj Alyaqobi and N. Adnan Bin Yahaya, “A Systematic Review on Image Data Protection Methods,” *Int. J. Informatics Inf. Syst.,* vol. 5, no. 3, pp. 131–141, Sep. 2022

[23] T. McIntosh, A. S. M. Kayes, Y.-P. P. Chen, A. Ng, and P. Watters, “Applying staged event-driven access control to combat ransomware,” *Computers & Security*, vol. 128, no. 1, pp. 103160–103167, 2023. doi:10.1016/j.cose.2023.103160