IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204169 |

Ransomware Detection System using Machine Learning Algorithms

Arshi Bhaldar [1], Vibha Shinde[2], Siddhanth Das[3], Gaurav Singh [4], Abhay Patil [5]

U.G. Student, Department of Information Technology, Rajiv Gandhi Institute of Technology (Mumbai University), Mumbai, Maharashtra, India[1,2,3,4]

Asst. Professor, Department of Information Technology, Rajiv Gandhi Institute of Technology (Mumbai University), Mumbai, Maharashtra, India[5]

ABSTRACT:Malicious attacks, malware, and ransomware families create significant cybersecurity concerns and have the power to cause significant damage to computer infrastructure, data centres, web pages and mobile applications in a variety of enterprises and organisations. Traditional anti-ransomware programmes have a hard time fending off freshly created, complex threats. Therefore, the development of innovative ransomware solutions can be successfully accomplished using contemporary techniques such as classical and neural network-based architectures. In this study, we provide a framework for feature selection-based ransomware identification and avoidance that classifies the security level using a range of machine learning methods, including neural network-based designs.

KEYWORDS:Ransomware, Machine Learning, Malware

I. INTRODUCTION

Cybersecurity is constantly under risk due to the constantly changing universe of harmful programmers and assaults, including ransomware and malware variants. These assaults have the potential to seriously impair computer networks, data centers, web-based applications, and mobile applications in a variety of enterprises and organizations. For instance, ransomware is made to prevent these types of assaults by preventing victims from accessing their computer data with impenetrable encryption techniques that can only be deciphered by an intruder. Nevertheless, eliminating ransomware can cost victims an awful lot of financial resources since their data may be compromised, and they frequently have to reimburse the perpetrator a ransom to recover control and keep their information confidential. Numerous kinds of ransomware exist, such as locker ransomware, which prevents accessibility to additional computers and gadgets, and crypto ransomware, which encrypts files along with additional data to prevent transmission. Events-based, statistical, and data-driven ransomware identification strategies are frequently unable to counter increasingly sophisticated assaults. In order to establish the best defenses and safety precautions against these hostile acts of violence, the investigation community must employ cutting-edge technology. The use of machine learning for ransomware detection is one interesting research area. Automated malware identification, including ransomware detection, can be improved by utilizing methods based on machine learning, such as dynamic behavioral analysis. This will strengthen safety measures contrary to emerging advanced threats. The creation of cutting-edge ransomware alternatives that make use of machine intelligence has enormous promise for thwarting the constantly changing terrain of noxious assaults.

II. EXISTING SYSTEM

There are two types of detection techniques commonly used by malware analysts in existing systems:

Static Detection:

In static detection, malware analysts analyse the disassembled code of an executable file without actually executing the binary file. Static file analysis can be used, for instance, to identify whether a file is suspicious when an alert about its possible maliciousness is raised on a crucial server within an organisation. Such an analysis searches for recognised fraudulent code patterns or dubious strings, for instance frequently utilised file extensions and frequent phrases found in ransom records. Without executing the code, static file analysis has the ability to determine whether an executable file is ransomware or any other sort of malware.

Dynamic Detection:

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204169 |

Dynamic detection involves monitoring the behaviour of malware by executing it in a controlled environment. Ransomware assaults, for case in point, can be identified in real-time and perhaps halted automatically depending on the safety measure that is implemented by watching the directory structure for bulk file actions, which include rename, write, or delete, throughout a specific time frame. In order to verify and authenticate files, File Integrity Monitoring (FIM) tools ought to be utilised. These tools compare files to a recognised, trustworthy "baseline," warning users whenever a file has been modified, revised, or compromised. This can aid in the early detection of ransomware attacks and limit impact.

Locker ransomware, which restricts accessibility to the computer or gadgets, and crypto ransomware, that encrypts files or additional information to prevent accessibility, are the two primary forms of ransomware assaults. It may prove difficult to recuperate against such assaults without covering the ransom. Event-based, statistical, and data-driven ransomware identification strategies can fail to be sufficient to counter increasingly advanced assaults. Therefore, leveraging both static and dynamic detection techniques, along with advanced tools like FIM and other innovative approaches, is crucial in effectively detecting and mitigating ransomware attacks.

III. PROPOSED SYSTEM

In order to identify and avoid ransomware, we have created an extensive structure that combines feature selection methods with several machine learning algorithms, such as neural network-based designs. We use machine learning methods like Random Forest (RF) and XG Boost to classify ransomware by applying them to a carefully chosen set of attributes. We have included feature selection techniques like variance inflation factor to find and remove poor variant and highly linked characteristics from the data in order to assure the efficiency of our system.

Innovative machine learning algorithms present great opportunities for developing cutting-edge identification and avoidance procedures in the sphere of ransomware alternatives. Our methodology makes use of these developments to precisely categorise the threat level of prospective ransomware attacks and to enable preventative actions to lessen their effects. Our system offers a reliable and effective method for identifying and avoiding ransomware and mitigation, advancing research in this field by merging feature selection strategies, machine learning algorithms, and neural network-based structures. With assaults becoming more frequent and severe, ransomware has grown to be an actual concern to companies as well as consumers. Numerous strategies have been put out to lessen the effects of ransomware attacks. Utilising machine learning to identify ransomware is one such approach.

The following steps are part of the machine learning-based suggested approach for ransomware detection:

1. **Information Gathering**: The initial step is to gather a dataset of ransomware and benign files. The dataset ought to include sufficient data samples for the machine learning model to be trained.
2. **Feature Extraction**: After collecting the data, features must be extracted. It is possible to extract attributes like file size, file type, entropy, and API calls. The machine learning model will be trained using these attributes.
3. **Machine Learning Framework**: A machine learning framework is trained using the attributes that were extracted. The prediction model can be trained using a variety of machine learning techniques, including Random Forest and XG Boost. The model needs to be trained on a sizable dataset of malicious and safe files.
4. **Evaluation**: To find out whether the model is accurate in spotting ransomware, it must be analysed after training.

Precision, recall, and F1-score are just a few of the metrics that can be used to evaluate something.

5. **Implementation**: The model can finally be used in a real-time setting to identify ransomware assaults. The system should be constantly checked for any suspicious activity, and if any is found, it should notify the user.

IV. METHODOLOGY

To detect ransomware in our research, we used decision tree classifier, random forest classifier, and XG Boost classifiers with classic machine learning classifiers. The structure of our model is shown in Figure 4.1. We standardised the ransomware data, which had different scales, to a same range to guarantee consistency amongst the variables. A subsection of essential characteristics within the dataset were also found and chosen using feature-selection techniques. In order to distinguish ransomware from valid observations, these chosen attributes were then fed into the various classifiers. The 138,047 samples and 54 attributes that made up the dataset for our investigation were taken via M. Mathur's work on "Ransomware (Malware) Detection using Machine-Learning." In these samples, ransomware was identified in 70% of the cases, while valid observations made up the remainder 30%. Our method aimed at precisely identifying ransomware in a blended dataset, advancing the field of ransomware detection research by combining

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204169 |

conventional machine learning classifiers and neural network-based architectures with standardised information and feature selection techniques.

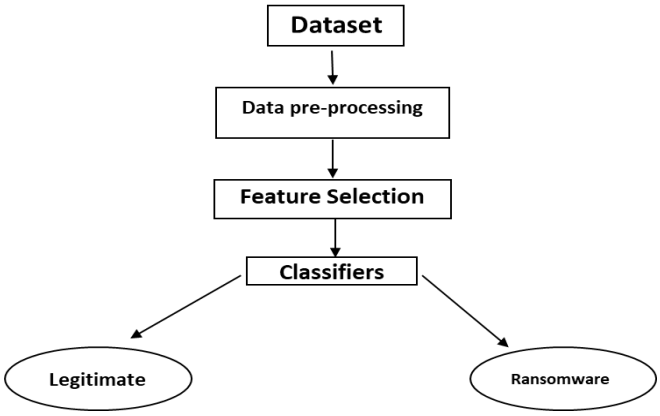

Figure 4.1:Framework to detect ransomware

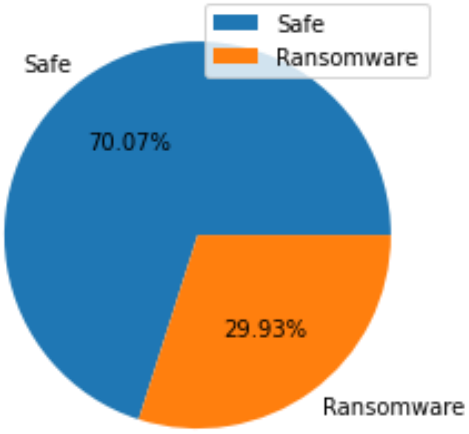

Figure 4.2: Distribution of dataset

V. EXPERIMENTAL RESULTS

The outcomes show that the Random Forest classifier conducted better than other classifiers by obtaining the greatest accuracy, F-beta, and precision scores with adequate regularity.

Results from Random Forest:

	Precision	Recall	F1-score	Support
0	0.99	0.99	0.99	31843
1	0.98	0.99	0.98	13713
Accuracy			0.99	45556
Micro Avg	0.99	0.99	0.99	45556
Weighted Avg	0.99	0.99	0.99	45556

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204169 |

Results from XG Boost:

	Precision	Recall	F1-score	Support
0	0.99	0.98	0.99	31843
1	0.95	0.99	0.97	13713
Accuracy			0.98	45556
Micro Avg	0.97	0.98	0.98	45556
Weighted Avg	0.98	0.98	0.98	45556

VI. FUTURE SCOPE

The emergence of malware, including ransomware, poses a persistent and severe threat to computer systems and networks, necessitating the development of effective cybersecurity measures. To mitigate the risks associated with ransomware attacks, it is imperative to implement automated systems that can provide robust security. In our system,we utilised a range of machine learning algorithms, such as neural network-based classifiers, in our approach to improve the identification of ransomware and categorization. Machine learning algorithms utilize statistical patterns to make accurate predictions and can be applied to diverse domains such as document analysis, fraud detection, KYC processing, and high-frequency trading. The future scope of machine learning in the banking sector is promising, as it has the potential to revolutionize and strengthen cybersecurity measures. By incorporating machine learning techniques into our ransomware detection system, we aim to enhance its effectiveness and contribute to the advancement of cybersecurity research and practices.

VII. CONCLUSION

Malware, particularly ransomware, poses serious security dangers to both individuals and organisations. The creation of automated systems for precise malware categorization and recognition is essential to reducing these threats and enhancing cybersecurity measures. In the present investigation, we used the Random Forest and XG Boost algorithms in a framework based on feature selection to effectively classify and identify ransomware. We conducted multiple trials to assess the efficacy of the models using an arsenal of ransomware samples. The experimental findings showed that the Random Forest classifier conducted better than other classifiers, consistently attaining the greatest accuracy, F-beta, and precision scores. These outcomes demonstrate how well our conceptual framework works to improve ransomware identification and avoidance strategies.

REFERENCES

[1] M. Masum, M. J. Hossain Faruk, H. Shahriar, K. Qian, D. Lo and M. I. Adnan, "Ransomware Classification and Detection With Machine Learning Algorithms," 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), 2022, pp. 0316-0322, doi: 10.1109/CCWC54503.2022.9720869.

[2] U. Adamu and I. Awan, “Ransomware prediction using supervised learning algorithms,” Proc. - 2019 Int. Conf. Futur. Internet Things Cloud, FiCloud 2019, pp. 57–63, 2019, doi: 10.1109/FiCloud.2019.00016.

[3] F. Noorbehbahani and M. Saberi, “Ransomware Detection with Semi-Supervised Learning,” 2020 10h Int. Conf. Comput. Knowl. Eng. ICCKE 2020, pp. 24–29, 2020, doi: 10.1109/ICCKE50421.2020.9303689.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details