NEXT-GEN RANSOMWARE DEFENSE: MACHINE LEARNING- POWERED THREAT DETECTION

Sathya S[1*], Bhuvaneshwari S[2], Subalakshmi M[3], Vidhya S[4]

[1]Assistant Professor, Department of Computer Science and Engineering,*
[2,3,4]Final Year UG Scholars, Department of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore, India

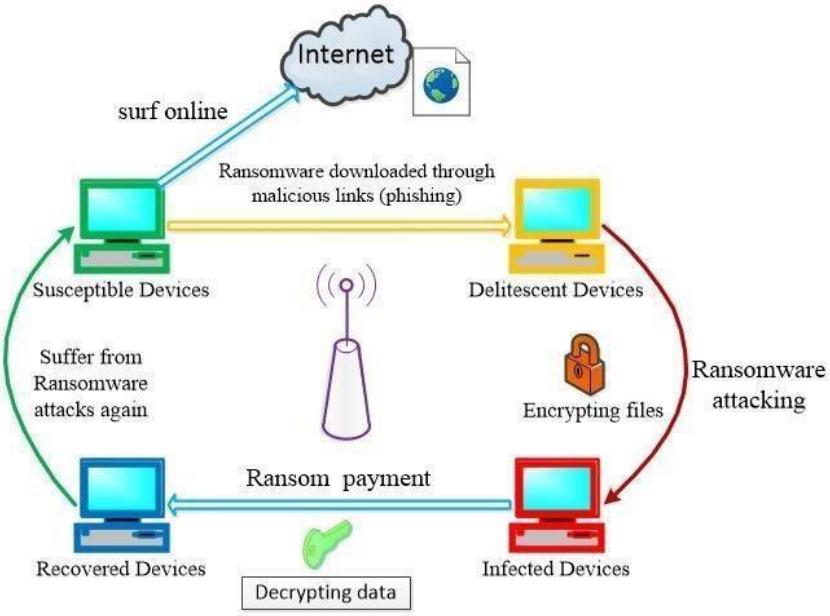
**Corresponding author:*

ABSTRACT
Ransomware attacks pose a significant and escalating threat to individuals and organizations, causing substantial financial and operational disruptions. This research explores the application of machine learning techniques for the proactive prediction of ransomware activity. By analyzing a comprehensive dataset of system behaviors, network traffic, and file system modifications, we develop predictive models capable of identifying potential ransom ware attacks before encryption occurs. We employ a range of machine learning algorithms, including Random Forest, Lazy Predict, to classify malicious activity. Our methodology incorporates feature engineering to extract relevant indicators of ransom ware behavior, and we evaluate the performance of our models using rigorous testing and validation of datasets using LIME. In order to improve the detection rate of ransomware, the data imbalance was addressed by utilizing the SMOTE- Tomek method, which allowed for a more robust machine learning prediction model. The results demonstrate the effectiveness of machine learning in enhancing ransomware detection and mitigation, offering a valuable tool for strengthening cyber security defenses.

Keywords*: Ransom ware attacks, Machine Learning, Predictive models, Random Forest, LazyPredict, Feature engineering, LIME, Data imbalance, SMOTE-Tomek, Detection rate, Cyber Security.*

1. INTRODUCTION

Ransomware attacks represent a significant and evolving threat in the digital landscape. These attacks involve malicious software that encrypts a victim's data, effectively locking them out of their own files and systems. Cybercriminals then demand a ransom, typically in crypto currency, in exchange for the decryption key. This form of cyber extortion can cripple individuals, businesses, and even critical infrastructure, leading to substantial financial losses and operational disruptions. Ransomware was originally introduced to target individual systems such as the personal computers of ordinary citizens. Nevertheless, attackers realized its full potential when they began to target organizations that were willing to pay to retrieve and protect their employee and customer data. Nowadays, most ransomware attacks happen to businesses and other organizations, including small and medium organizations that lack the resources to fully shield themselves from such attacks. Some of the most affected industries include banking, utilities, education, government, and manufacturing. Attackers also execute more ransom ware attacks on affluent regions and countries of the world, like the United States, Canada, and Australia. They do this for 2 reasons: wanting to extort more money from rich companies situated in these regions, and these regions tending to have higher PC adoption rates.



In the realm of ransom ware detection, the Random Forest algorithm emerges as a robust and adaptable tool, particularly effective in navigating the complexities inherent in modern cyber threats. As an integral component of a comprehensive defense strategy, this machine learning technique excels at discerning subtle patterns within vast datasets of system behavior. The process typically begins with meticulous feature extraction, where critical indicators like file system modifications, network communication anomalies, and resource utilization spikes are transformed into quantifiable attributes.

Subsequently, the Random Forest algorithm constructs an ensemble of decision trees, each trained on randomized subsets of data and features, mitigating over fitting and enhancing generalization. This ensemble approach allows for a collective decision-making process, where the majority vote of the trees determines the classification of a given activity as either benign or malicious. Its inherent resilience to noisy data, a common characteristic of cyber security logs, and its ability to identify key predictive features, render it invaluable. By integrating Random Forest into end point detection and response (EDR) and network intrusion detection systems (NIDS), organizations can bolster their defenses, enabling real-time threat identification and mitigation, thereby minimizing the impact of these increasingly sophisticated attacks.

OBJECTIVE

The primary objectives of this study is to proactively identify and neutralize ransomware activity before it can significantly encrypt data and disrupt operations and even accurately and efficiently identify malicious ransomware activity in real-time or near real-time, thereby preventing or mitigating its impact on systems and data. The foremost aim is to design a method to assess these properties to identify potential malware. The script should flag any **.exe** files that exhibit unusual or suspicious values for these properties, indicating a higher likelihood of being malicious. We are developing a script for an antivirus package aimed at identifying malware before executing .exe files on Windows. Our approach involves analyzing specific properties of executable files, such as: ImageBase, VersionInformationSize and SectionsMaxTrophy. To achieve higher prediction accuracy and reliability, the system combines the robustness of the Random Forest algorithm with the efficiency of automated machine learning (Auto ML) tools like Lazy Predict to enhance ransomware prediction. Initially, Lazy Predict rapidly evaluates numerous machine learning models on extracted system behavior features, providing a quick assessment of their performance. This automated exploration identifies Random Forest as a potentially high-performing model for this specific task, given its ability to handle

complex, high-dimensional datasets and its inherent resistance to over fitting. Following this initial assessment, Random Forest is then fine-tuned for optimal performance. This involves hyper parameter optimization, often achieved through techniques like grid search or Bayesian optimization, to maximize accuracy and minimize false positives. The algorithm analyzes features such as file system changes, network traffic anomalies, and resource utilization patterns, learning to distinguish between benign and malicious activities. This combination of Auto ML's rapid model assessment and Random Forest's robust predictive capabilities allows for the development of a highly effective ransom ware prediction system, enabling real-time or near real-time detection and mitigation of threats. The system learns to adapt to evolving ransom ware variants, ensuring continuous improvement in its ability to protect against these dynamic Cyber threats. By using Lazy Predict, ransom ware prediction models can be developed faster and more efficiently, reducing the need for manual model tuning and allowing practitioners to focus on fine-tuning other parts of the security system.
Additionally, the integration of Random Forest with Lazy Predict enables rapid testing of several algorithms (e.g., SVM, Logistic Regression, etc.) on ransom ware-related datasets, allowing for a more comprehensive model selection process. The combination of these techniques can significantly enhance the accuracy and speed of ransom ware detection systems, making them more effective in real-time cyber security scenarios. Moreover, this approach can also be extended to continuously monitor systems and adapt to evolving ransomware tactics, ensuring that the security measures stay up-to-date.

METHODOLOGY

To achieve the goal of accurate and efficient ransomware prediction, this study implements multiple machine learning algorithms, each offering unique mechanisms to analyze the datasets. Ransom ware prediction using machine learning, particularly random forest, employs several strategies to enhance cyber security. These approaches primarily involve analyzing system behaviors, such as file modifications, network traffic, and resource usage, to detect anomalies indicative of ransom ware activity.

Dataset:

The dataset used in this project comprises about 62,000+ samples from the Kaggle dataset that includes both benign and malicious software. The dataset consists of PE file characteristics extracted from a collection of Windows executables and DLL files. Each entry represents a unique file with various attributes extracted from its PE header and structure. The dataset includes both benign software samples and known malware samples (identified by VirusShare hashes). The dataset is designed for malware detection and analysis research, containing various static features extracted from PE file headers and structures.

Feature Extraction

Feature selection is a critical step in this research, aimed at identifying the most predictive features to enhance the accuracy and efficiency of the machine learning models. By focusing on the most relevant features, we can reduce the complexity of the model, improve performance, and avoid over fitting.
We employed IV (Information Value) and WoE (Weight of Evidence) to predict the relevant features used for the algorithm because they offer a robust and statistically sounds way to assess the predictive power of variables, particularly in binary classification problems. IV (Information Value) and WoE (Weight of Evidence) are employed to identify the most relevant features for prediction by transforming categorical or continuous variables into a scale reflecting their predictive power, particularly in binary classification. WoE quantifies the "weight" of evidence each feature value provides in distinguishing between classes, while IV summarizes this across all feature categories, yielding a single value representing predictive strength. These techniques effectively linearize relationships, handle missing values, and are robust to outliers, making them ideal for feature selection and ranking. By highlighting the most informative variables, IV and WoE enhance model accuracy and interpretability, ensuring that predictive models are built upon the most significant data attributes. In ransomware prediction, Information Value (IV) and Weight of Evidence (WoE) are crucial for pinpointing the most impactful system behavior features. WoE transforms complex system metrics, like file modification patterns or network traffic anomalies, into a predictive scale, highlighting which behaviors strongly correlate with ransom ware activity. IV then ranks these transformed features, revealing which contribute most significantly to distinguishing malicious from benign actions. By quantifying the predictive power of diverse system attributes, IV and WoE enable security analysts to prioritize critical indicators, enhancing the accuracy and efficiency of ransom ware detection models.

Random Forest

The Random Forest algorithm, a powerful ensemble learning technique, constructs multiple decision trees during training by employing bootstrap aggregating (bagging) and random feature selection. Each tree, trained on a random subset of the data and features, contributes a vote for classification or an average for regression. This approach minimizes over fitting and enhances generalization, resulting in high accuracy and robustness to noisy data. Its ability to handle high- dimensional datasets and provide feature importance metrics makes it a versatile tool for various applications, from image recognition and fraud detection to stock price prediction. In ransomware prediction, the Random Forest algorithm excels by analyzing complex system behaviors to distinguish malicious activity from benign operations. By constructing an ensemble of decision trees, each trained on randomized subsets of system features like

file modifications, network traffic, and resource usage; it effectively mitigates over fitting and enhances prediction accuracy. This robust approach allows for the identification of subtle patterns indicative of ransomware, even in noisy environments, proving invaluable for real-time threat detection within endpoint security systems.

Logistic Regression:
Logistic regression is a statistical method used for binary classification, predicting the probability of a binary outcome (e.g.., yes/no, true/false). Unlike linear regression, which predicts continuous values, logistic regression models the relationship between independent variables and the probability of the dependent variable using a logistic function (sigmoid), which constrains the output between 0 and 1. It estimates coefficients that indicate the impact of each independent variable on the log- odds of the outcome, making it suitable for scenarios where the target variable is categorical. In ransom ware prediction, logistic regression can be employed to classify system activities as either benign or malicious by analyzing various system behaviors. By modeling the probability of an activity being ransom ware based on features like file system changes, network traffic patterns, and resource usage, logistic regression identifies patterns indicative of malicious encryption processes. While perhaps less complex than other machine learning algorithms like Random Forest, logistic regression's interpretability allows security analysts to understand the influence of specific features on ransomware likelihood. This method offers a computationally efficient approach to real-time or near real-time detection, enabling timely intervention and mitigation of ransomware threats, particularly in scenarios where computational resources are limited.

Support Vector Machine (SVM):
Support Vector Machines (SVMs) are powerful supervised learning algorithms used for classification and regression. They operate by finding an optimal hyper plane that maximally separates data points of different classes in a high-dimensional space. For non-linearly separable data, SVMs employ kernel functions to map the data into a higher-dimensional space where linear separation is possible. The algorithm focuses on identifying support vectors, which are the data points closest to the hyper plane, as they determine the margin and thus the decision boundary. Support Vector Machines (SVMs) can be effectively utilized for ransomware prediction by classifying system behaviors as either benign or malicious. By analyzing features such as file system modifications, network traffic patterns, and CPU usage, SVMs can identify complex, non-linear patterns indicative of ransomware activity. SVMs focus on identifying support vectors, which are crucial for defining the decision boundary, making them robust against outliers and effective in high-dimensional feature spaces commonly encountered in cyber security data. This approach enables the development of accurate and efficient ransomware detection systems, capable of identifying subtle malicious behaviors and facilitating timely intervention to prevent data encryption and system disruption.

Naïve – Bayes:
Naive Bayes is a probabilistic machine learning algorithm used for classification tasks. It's based on Bayes' theorem, assuming that features are conditionally independent given the class label, hence the "naive" aspect. Despite this simplifying assumption, it often performs surprisingly well in real-world applications. Naive Bayes calculates the probability of a data point belonging to a particular class by multiplying the probabilities of its individual features occurring in that class. It's particularly effective for text classification, spam filtering, and other applications with high-dimensional data, due to its computational efficiency and ability to handle large datasets. There are different variants of Naive Bayes, such as Gaussian, Multinomial, and Bernoulli, each suited for different types of data distributions. Naive Bayes can be applied to ransomware prediction by calculating the probability of system behaviors indicating malicious activity. By assuming independence between features like file modifications, network connections, and resource usage, the algorithm efficiently classifies activity as either benign or ransomware. Despite its simplifying assumptions, Naive Bayes can be effective in handling high- dimensional cyber security data, providing a computationally efficient method for real-time or near real-time detection. It calculates the likelihood of an activity being ransomware based on the observed feature patterns, enabling rapid assessment and potential intervention. While perhaps less complex than other algorithms, its speed and ability to handle large datasets make it a viable option for ransomware detection, especially in environments where computational resources are limited.

LazyPredict – AutoML:
LazyPredict is an automated machine learning (AutoML) library designed to streamline the model selection process. It rapidly trains and evaluates numerous machine learning models on a given dataset, providing a quick overview of their performance metrics. This allows users to efficiently identify promising models, such as Random Forest, for further fine-tuning and optimization. By automating the initial stages of model exploration, LazyPredict significantly reduces the time and effort required to build predictive models, making it a valuable tool for quickly assessing the potential of various algorithms and identifying the most suitable candidates for a specific task. In the context of ransom ware prediction, LazyPredict serves as a rapid prototyping tool, quickly evaluating a wide array of machine learning models against extracted system behavior features. This allows security analysts to efficiently identify which algorithms, including Random Forest, demonstrate the highest potential for accurately distinguishing malicious ransom ware activity from benign operations. By automating the initial model selection phase, LazyPredict accelerates the development of effective prediction systems, enabling faster iteration and optimization. This rapid assessment helps to

pinpoint the most promising models for further fine-tuning, ultimately enhancing the speed and accuracy of ransom ware detection and response.

LIME:
LIME, or Local Interpretable Model-agnostic Explanations, plays a crucial role in dataset testing by providing insights into individual machine learning predictions. It illuminates the features that most influence a model's output for specific data points, revealing potential biases, anomalies, or errors within the dataset. By highlighting sensitive or irrelevant features, LIME aids in ensuring fairness and identifying data quality issues that might compromise model reliability. This tool effectively validates model behavior on individual instances, enhancing trust and facilitating debugging, ultimately improving the trustworthiness and robustness of machine learning applications by pinpointing data-related issues. In ransom ware prediction, LIME is instrumental for dataset testing by providing localized explanations for individual predictions, revealing which system behavior features— like file modifications, network traffic, or resource usage— contributed most significantly to a model's ransom ware classification. This allows security analysts to validate if the model is relying on relevant indicators or if it's being influenced by dataset anomalies or biases. By highlighting critical features, LIME aids in identifying potential data quality issues that could lead to false positives or negatives, enhancing their liability of ransomware detection systems. This localized interpretability fosters trust in the model's decisions, enabling more informed security responses and improving the overall robustness of the predictive system against evolving ransom ware threats.

SMOTE-TOMEK:
SMOTE-TOMEK is a powerful resampling technique designed to address data imbalance, a common challenge in machine learning datasets where one class significantly outnumbers the others. It combines the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class, with the Tomek links method, which removes ambiguous or noisy samples from the overlapping regions between classes. SMOTE- TOMEK effectively balances the class distribution while simultaneously cleaning the dataset, leading to improved model performance, particularly in classification tasks where the minority class is crucial, such as fraud detection or ransomware prediction. In ransomware prediction, where malicious samples are often significantly outnumbered by benign system behaviors, SMOTE-TOMEK becomes a crucial preprocessing step to mitigate data imbalance. By synthesizing new ransom ware samples using SMOTE and subsequently removing ambiguous data points with Tomek links, it effectively balances the dataset, enabling machine learning models, like Random Forest, to learn more robustly from the minority class. This technique ensures that the model isn't biased towards the majority of benign data, thereby improving its ability to accurately detect ransomware threats. By addressing the data imbalance, SMOTE-TOMEK enhances the model's sensitivity and specificity, crucial for minimizing false negatives and false positives in real-time threat detection systems, ultimately leading to more effective proactive defense against ransom ware attacks.

2. PROBLEM DEFINITION

The core challenge in ransomware prediction lies in developing a reliable and accurate system that can proactively identify malicious activity before significant data encryption occurs. This necessitates the creation of a model capable of discerning subtle patterns within complex system behaviors, such as file modifications, network traffic, and resource utilization, to distinguish ransomware from benign operations. The system must be robust against evolving ransom ware variants, minimize false positives to avoid operational disruptions, and address data imbalances that can skew prediction accuracy. Ultimately, the objective is to build a predictive model that enables timely intervention and mitigation, minimizing data loss and operational downtime in the face of increasingly sophisticated ransom ware attacks. The development of a robust ransomware prediction system faces significant challenges, primarily centered on achieving accurate and reliable detection while mitigating inherent risks. Overfitting poses a critical problem, where models trained on limited datasets may fail to generalize to unseen ransom ware variants, leading to poor real-world performance. Anomalies, such as rare but legitimate system behaviors, can be misclassified as malicious, resulting in false positives and operational disruptions. Furthermore, security issues, including adversarial attacks that manipulate system behavior or evade detection mechanisms, threaten the integrity of the predictive system. Thus, the objective is to create a predictive model that balances sensitivity and specificity, minimizes over fitting, handles anomalies effectively, and incorporates robust security measures to ensure reliable and trustworthy ransom ware detection in dynamic threat landscapes. The challenge lies in developing a robust and efficient malware detection script for a Windows antivirus package, focusing on the pre-execution analysis of .exe files. Thus the script should be designed to minimize false positives while maximizing the detection rate of actual malware, requiring a thorough understanding of typical benign executable characteristics and the ability to discern deviations that strongly correlate with malicious intent. This task demands a balance between performance and accuracy, ensuring that the antivirus package can effectively protect users against evolving malware threats by proactively analyzing executable properties before execution.

EXISTING SYSTEM
In the current cyber security landscape, ransom ware prediction, and detection primarily rely on signature-based methods and static/dynamic analysis. Signature-based detection identifies known ransom ware variants by matching file hashes or patterns to a database of known malware signatures, offering quick

detection of recognized threats. Static analysis examines file structures and code without execution, while dynamic analysis executes files in a controlled environment (sandboxes) to observe behavior. These methods, though historically prevalent, struggle with zero-day attacks and polymorphic ransom ware, which can evade signature matching. Dynamic analysis, while effective, is resource-intensive and can be bypassed by malware designed to detect sandbox environments. These systems often lack the adaptability to address the rapidly evolving nature of ransomware, resulting in delayed responses and potential data breaches. The traditional ransom ware detection systems, while historically foundational, suffer from several critical disadvantages. Signature-based detection is inherently limited to recognizing known ransom ware, rendering it ineffective against novel or polymorphic variants that alter their code to evade detection. Static analysis, though faster, struggles with obfuscated or packed malware, while dynamic analysis, despite its behavioral insights, is resource-intensive and vulnerable to sandbox evasion techniques. Furthermore, these systems often operate reactively; analyzing files post-execution, which results in delayed responses and potential data encryption before detection. The lack of adaptability to the rapidly evolving ransom ware landscape, coupled with the potential for high false-positive rates due to the complexity of modern malware, hinders their ability to provide comprehensive and proactive protection.

DISADVANTAGES:

- Limited adaptability: Cannot handle new or modified ransomware.
- Heuristic approach: High false positives (normal applications may be misclassified).
- Poor scalability: Inefficient for large datasets and multiple file types.
- Sandbox analysis: Slow detection speed, leading to delayed action.
- Network Analysis: Relies on constant updates to signature databases.
- Static analysis: Struggles with obfuscated/packed malware.

3. PROPOSEDSYSTEM

This paper proposes a proactive ransomware prediction system leveraging machine learning to mitigate the escalating threat of ransom ware attacks. By analyzing comprehensive system behavior, network traffic, and file modification datasets, the system employs algorithms like Random Forest and LazyPredict to classify malicious activity before encryption. Feature engineering extracts key indicators, and model performance is rigorously validated using LIME to ensure interpretability. [2]To address data imbalance, SMOTE-Tomek is utilized, enhancing the model's robustness. The system aims to provide a valuable tool for strengthening cyber security defenses by enabling early detection and mitigation of ransomware threats, demonstrating the effectiveness of machine learning in proactive ransom ware prediction.

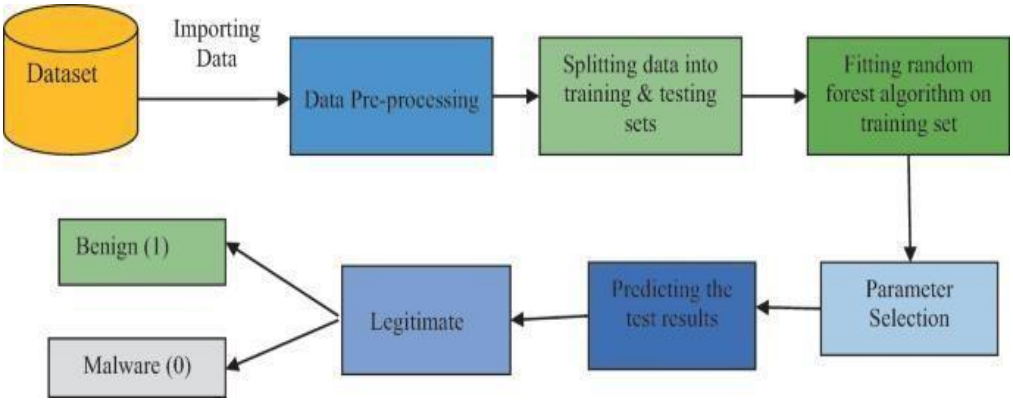
ADVANTAGES:

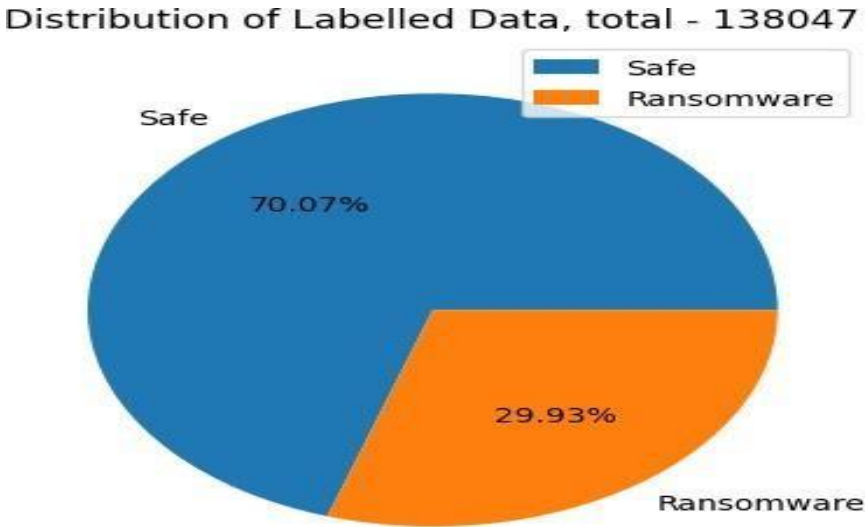
- Accuracy improvement: Ensemble model achieves 97.70% accuracy.
- More accurate than traditional rule-based systems.
- Detects unknown (zero-day) ransom ware.
- Faster response time, preventing data encryption.
- Less human effort (automated model selection + training).
- Explainable AI (LIME) helps security teams trust model decisions

4.3 IMPLEMENTATION OF THE MODEL:

The diagram illustrates the implementation of a machine learning model, specifically using the Random Forest algorithm, for a binary classification task, likely malware detection. The process begins with importing a dataset which is then subjected to data pre-processing to clean and prepare it for analysis. Subsequently, the data is split into training and testing sets, ensuring the model's performance can be evaluated on unseen data. The core of the process involves fitting the Random Forest algorithm on the training set**,** allowing the model to learn patterns that distinguish between the two classes (likely "Benign" and "Malware"). After training, the model's performance is optimized through parameter selection**.** Finally, the trained model is used to predict the test results**,** evaluating its accuracy on the unseen test data. The diagram also indicates a step where the classes are labeled as "Legitimate" (presumably synonymous with "Benign") and "Malware", suggesting a focus on distinguishing between safe and malicious files. This workflow outlines a standard machine learning pipeline for binary classification, with a focus on Random Forest and meticulous data handling. The implementation process began with the acquisition of a comprehensive dataset from Kaggle, encompassing over 62,000 samples of both benign and malware software. This dataset underwent rigorous preprocessing, involving data cleaning and transformation to ensure quality and compatibility for machine learning algorithms. Subsequently, Information Value (IV) and Weight of Evidence (WoE) were employed to identify and rank the most relevant features, streamlining the dataset for optimal predictive performance. The dataset was then partitioned into training and testing sets to facilitate model development and evaluation. LazyPredict, an AutoML library, was utilized to rapidly assess the performance of various machine learning models, and the results were evaluated using a confusion matrix to gauge accuracy and effectiveness. Following this, the LIME algorithm was applied to provide localized explanations for individual predictions, shedding light on the features driving classifications of both legitimate and ransomware files. To

address the inherent data imbalance, the SMOTE-TOMEK technique was implemented, generating synthetic samples and removing noisy data points to create a more balanced dataset. Finally, the entire evaluation process, including LIME explanations and confusion matrix analysis, was repeated on the balanced dataset, allowing for a comparative assessment of the impact of SMOTE-TOMEK on model performance and interpretability. This comprehensive approach ensured a robust and transparent ransomware prediction system, addressing both performance and interpretability concerns.





5. CONCLUSION AND FUTURE WORK

CONCLUSION:

The proposed system for the ransomware prediction, as outlined in the diagram and description, presents a well-structured approach to building a robust machine learning model for binary classification, likely malware or ransom ware detection. The process begins with importing a comprehensive dataset, followed by meticulous data pre processing to ensure data quality. Feature selection, utilizing IV and WoE, streamlines the model by focusing on the most relevant attributes. The use of LazyPredict for automated model selection accelerates the initial stages, allowing for rapid evaluation and selection of the most promising algorithm, likely Random Forest given the diagram. The evaluation of the confusion matrix provides critical insights into the model's performance, while LIME explanations enhance interpretability, shedding light on the model's decision-making process. Addressing data imbalance with SMOTE-TOMEK further refines the model, improving its ability to accurately classify minority class instances. Repeating the evaluation post SMOTE-TOMEK ensures a thorough assessment of its impact, highlighting the importance of data balancing in achieving optimal model performance. This comprehensive approach, combining automated model selection, interpretability analysis, and data balancing, results in a highly effective and transparent classification system, capable of accurately distinguishing between benign and malicious files.

FUTURE WORK:

Although the current system shows high accuracy and efficiency, several improvements are planned to enhance its performance and usability. Future work in ransom ware prediction should focus on enhancing real-time detection and adaptability to evolving threats. This includes developing more sophisticated feature engineering techniques to capture subtle behavioral changes indicative of new ransom ware variants. Incorporating advanced deep learning models, such as recurrent neural networks (RNNs) or transformers, could improve the detection of complex attack patterns and long-term dependencies in system behavior. Research should also explore federated learning to enable collaborative model training across multiple endpoints without centralizing sensitive data, addressing privacy concerns. Furthermore,

integrating threat intelligence feeds and developing explainable AI methods to provide actionable insights into prediction rationales will be crucial for proactive defense strategies. Finally, focusing on adversarial machine learning techniques to fortify models against evasion attempts will ensure the robustness of ransom ware prediction systems in dynamic cyber threat environments.

REFERENCES:

1. M. Benmalek, Ransomware on cyber-physical systems: taxonomies, case studies, security gaps, and open challenges Internet Things Cyber-Phys. Syst., 4 (2024), pp. 186-202,10.1016/j.iotcps.2023.12.001
2. M.Cen, F.Jiang, X.Qin, Q.Jiang, R.Doss, Ransomware early detection: a survey, Comput. Netw., 239 (2024), Article 110138, 10.1016/j.comnet.2023.110138
3. A.S.M.Al-Ruwili, A.M.Mostafa, Analysis of ransomware impact on android systems using Machine Learning techniques, Int. J. Adv. Comput. Sci. Appl.,14(11)(2023),10.14569/IJACSA.2023.0141178
4. A. Kapoor, A. Gupta, R. Gupta, S. Tanwar, G. Sharma, I.E. Davidson Ransomware detection,avoidance, and mitigation scheme: a review and future directions, Sustainability.,14(1)(2021),p.8,10.3390/su14010008
5. Md.A. Hossain, Md.S. Islam, A novel hybrid feature selection and ensemble-based machine learning approach for botnet detection Sci.Rep.,13(1)(2023),p.21207,10.1038/s41598-023- 48230-1
6. J.W. Hu, Y. Zhang, Y.P. Cui, Research on android ransomware protection technology, J. Phys. Conf. Ser.,1584(1)(2020),Article012004,10.1088/1742- 6596/1584/1/012004
7. A. Pagan, K. Elleithy, A multi-layered defense approach to safeguard against ransomware, 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, NV, USA (2021), pp. 0942-0947,10.1109/CCWC51732.2021.9375988
8. M.A.Hossain, M.S.Islam, Ensuring network security with a robust intrusion detection system using ensemble-based machine learning, Array (2023), Article100306,10.1016/j.array.2023.100306
9. A.Vehabovic, N.Ghani, E.Bou- Harb, J. Crichigno, A. Yayimli Ransomware detection and classification strategies, 2022 IEEE International Black Sea Conference on Communications and Networking (Black Sea Com)(2022), pp.316324,10.1109/BlackSeaCom54372.2022.9858296
10. B.Yamany, M.S.Elsayed, A.D.Jurcut, N.Abdelbaki, M.A. Azer, A holistic approach to ransomware classification: leveraging static and dynamic analysis with visualization Information, 15 (1) (2024), p. 46, 10.3390/info15010046
11. N. Singh, S. Tripathy, It's too late if exhilarate: early stage Android ransomware detection,Comput.Secur.,141(2024),Article103819,10.1016/j.cose.2024.103819
12. A. Moshood Abiola, M. Fadzli Marhusin, Signature- based malware detection using sequences of N-grams Int. J. Eng. Technol., 7 (4) (2018), p. 4,10.14419/ijet.v7i4.15.21432
13. C.Beaman, A.Barkworth, T.D.Akande, S.Hakak,M.K. Khan, Ransomware: recent advances, analysis, challenges and future research directions, Comput.Secur.,111(2021), Article102490,10.1016/j.cose.2021.102490
14. C.Zheng,*etal.*, SmartDroid: an automatic system for revealing UI-based trigger conditions in android applications Proceedings of the second ACM workshop on Security and privacy in smart phones and mobile devices, Association for Computing Machinery, New York, NY, USA (2012), pp. 93- 104,10.1145/2381934.2381950
15. V.Rastogi, Y.Chen, X.Jiang, Catch me if you can: evaluating android anti-malware against transformation attacks IEEETrans.Inf.ForensicsSecur.,9(1)(2014),pp.99- 108,10.1109/TIFS.2013.2290431
16. G. Canfora, E. Medvet, F. Mercaldo, C.A. Visaggio Detecting Android malware using sequences of system calls, Proceedings of the 3rd International Workshop on Software Development Lifecycle for Mobile, ACM, Bergamo Italy (2015), pp.13- 20, 10.1145/2804345.2804349
17. M.Bi, J.Xu, M.Wang, F.Zhou, Anomaly detection model of user behavior based on principal component analysis,J.AmbientIntell.Humaniz.Comput.,7(4)(2016),pp.547-554,10.1007/s12652-015- 0341-4
18. Y. Kaya et al., “Demystifying behavior-based malware detection at endpoints,” May 09, 2024, *arXiv*: 2405.06124. Accessed: Jul. 27, 2024. [Online]. Available: http://arxiv.org/abs/2405.06124
19. J. Yang, Z. Zhang,H. Zhang,J. Fan, Android malware detection method based on highly distinguish able static features and DenseNet, PLoS. One, 17 (11) (2022), Article e0276332, 10.1371/journal.pone.0276332