Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

MACHINE LEARNING-BASED TECHNIQUES FOR EFFECTIVE RANSOMWARE DETECTION

[1]*Kottha Vinay*, *MCA Student, Department of MCA*
[2] *Dr.K.Pavan Kumar, Ph.D, Associate Professor, Department of MCA*
[12]*Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool*

ABSTRACT
The need of effective and dynamic detection and mitigation techniques is growing in light of the increasing frequency and sophistication of ransomware assaults. Often, traditional mark-based methods are inadequate in identifying new and evolving ransomware variants. In order to increase the accuracy and adaptability of detection tools, this study explores the use of machine learning techniques for ransomware detection. It offers a thorough analysis of several machine learning techniques and algorithms, evaluating how well they detect ransomware trends. The findings provide crucial information for creating cybersecurity plans that are more robust and proactive in addressing the ever-changing ransomware threat scenario.

I. INTRODUCTION

The kind and complexity of assaults have significantly changed as a result of the constant growth of digital threats inside the computerised world [1]. Ransomware has emerged as a particularly formidable adversary among the plethora of digital dangers, causing unending disruptions to individuals, organisations, and fundamental systems [2]. By exploiting flaws in cybersecurity defences, ransomware encrypts sensitive data and demands excessive payments to be delivered. Although traditional mark-based detection tools show some promise, they often struggle to keep up with the rapid growth of ransomware variants that use cunning tactics to attempt to avoid detection [3]. As might be inferred from the test's ominous concept, integrating machine learning processes into cybersecurity frameworks has received significant attention [4-5]. Machine learning has the possibility for flexibility and proactive threat identification by enabling frameworks to learn and adapt via verified data and emerging trends [6-7]. The application of various machine learning techniques to improve the precision and sufficiency of detection components is the explicit focus of this paper's investigation into the field of ransomware detection. The next sections will carry out a thorough examination of several machine learning techniques for ransomware detection. We will examine the advantages and disadvantages of deep learning models, supervised learning algorithms, and unsupervised learning techniques [9]. The study will also provide a comprehensive methodology for integrating machine learning into existing cybersecurity systems while taking into account the real-world deployment problems. It is crucial to emphasise the critical role machine learning plays in improving the versatility and efficacy of ransomware detection systems as we set out on this investigation. Our goal is to use the power of intelligent algorithms to anticipate and prevent the changing tactics used by hackers in addition to identifying and stopping current ransomware attacks. In the never-ending game of cat and mouse, we hope that our study will provide insightful information that will help cybersecurity experts keep one step ahead of ransomware attackers [10]. The article is organised into the following sections: part II

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

clearly outlines the primary goals, while the introduction provides background for the study. Section III of a thorough literature review evaluates previous research critically, pointing out any gaps and placing the study in the larger academic context. Section IV The paper's main focus is on examining several machine learning techniques for ransomware detection. While the next section VI on data preparation outlines the procedures used to select and organise the dataset for analysis, Section V on methodology describes the study strategy and data gathering process. The corresponding portion of portion VII analyses the results, while Section VIII discusses the findings in light of the goals of the study and the body of current literature. Key contributions are included in Section IX, the conclusion, which also offers suggestions for more study. From the theoretical underpinnings to the actual application and analysis, this methodical approach guarantees a comprehensive investigation of machine learning in ransomware detection.

1.2 LITERATURE SURVEY

Abdullah, A., Kok, S. H., and Jhanjhi, N. Z. (2022). Pre-encryption detection algorithms are used to identify crypto-ransomware early. Computer and Information Sciences Journal of King Saud University, 34(5), 1984-1999.

Malware that uses an encryption technique to lock its victim's file for ransom is known as crypto ransomware. Due to a number of successful global assaults, its popularity within the cyber security community has increased at an alarming pace. Even if the victim decides to pay the ransom, the encryption utilised has permanently corrupted their digital data. As a result, this study suggests the Pre-Encryption Detection Algorithm (PEDA), which is capable of identifying crypto-ransomware before encryption is applied. PEDA has two detection levels. The first level was performed by comparing the ransomware's signature with that of a known crypto-ransomware before the ransomware could be triggered. The Secure Hashing Algorithm (SHA-256) was used to produce the signature, enabling quick and precise file content comparison. Learning Algorithms (LA), which may identify crypto-ransomware based on pre-encryption application program interfaces (API), were used in the second stage of detection. Based on 80:20 training and testing ratios, the LA generated a 100% recall rate; a 10-fold cross-verification test yielded a 99.9% recall rate. Furthermore, our study successfully discovered fourteen key APIs that may distinguish between goodware and ransomware. After that, it asks the victim to pay a ransom in order to release the resources. This kind of malware gained widespread recognition in the field of cyber security when Wannacry launched a global cyberattack in 2017 that infected over 200,000 systems in 150 different countries. Ransomware continues to be one of the most dangerous malwares that cyber security experts keep on high alert. Its influence on financial loss has grown despite a decrease in its infection incidence. This is because ransomware is being used by thieves to target corporations instead than the general public. As a result, the ransom demand is likewise much larger.

Ransomware is appealing to thieves for a number of reasons. The Internet's widespread use, which permits global connection, is the first factor (Diro et al., 2020, Almusaylim et al., 2020). While this greatly facilitates communication, it also increases the danger of cross-border cyberattacks. For instance, a cybercriminal may easily initiate an assault on a selected American firm from any location in Asia. Because of the physical distance and the disparate laws that apply to each nation, this transnational assault will provide challenges for the authorities. The widespread adoption of cryptocurrencies like the well-known Bitcoin is the second cause. The owner of this digital money may stay anonymous, which makes it more difficult for authorities to find them. As a

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

result, the ransom money may be obtained by the cybercriminal in a safe and secure manner.

The desire to save digital data is the third cause. Initiatives like "Go Paperless," "Reduce Your Carbon Footprint," "Digitisation," and others encourage the storing of data in digital format as opposed to hardcopy print. The fact that digital storage is becoming more affordable over time lends even more credence to this. Additionally, digital material has several benefits over physical print, like being searchable, taking up less space, being readily backed up or copied, and more. The encryption algorithm's maturity, accessibility, and significance make up the fourth component. In order to accomplish the security goals established by the security community for an information system, including confidentiality, integrity, availability, authenticity, and accountability (CIAAA), encryption is a crucial security technique. With the exception of availability, encryption aids in the accomplishment of all these goals. Ransomware, on the other hand, uses encryption to steal user data and hold it hostage.

The varieties of ransomware and encryption algorithms used by crypto-ransomware are shown in Fig. 1. One secret key is used for both encryption and decryption in symmetrical encryption. Although this kind of encryption performs better, it is not as strong. Two keys are used in asymmetrical encryption. For encryption, a single key—also known as the public key—is used. The decryption process uses a different key known as the private key. Although this kind of encryption is robust, it takes a long time to finish. The last kind, known as hybrid encryption, combines the two earlier forms of encryption to provide robust and high-performance encryption. This is accomplished by first utilising symmetrical encryption to encrypt the data or files, and then using asymmetrical encryption to encrypt the secret key. Ransomware is capable of using all three forms of encryption.

Kok, S. H., Jhanjhi, N. Z., Supramaniam, M., & Abdullah, A. (2019). A pre-encryption detection approach is used to prevent crypto-ransomware. 79 in Computers, 8(4).

The goal of ransomware, a relatively recent kind of intrusion assault, is to demand a ransom from its victim. Although ransomware attacks come in a variety of forms, the current study exclusively discusses crypto-ransomware as it renders data unrecoverable once the victim's files have been encrypted. Therefore, it was suggested in this study that machine learning be used to identify crypto-ransomware in the pre-encryption stage, or before it begins its encryption function. For the assault to be prevented from accomplishing its goal, successful detection at this point is essential. Important files and data may be backed up to a different place once the victim is aware that crypto-ransomware is there. Then, an effort can be made to remove the ransomware with the least amount of risk. As a result, we suggested a two-phase pre-encryption detection method (PEDA). A Windows application programming interface (API) produced by a questionable software would be recorded and examined using the learning algorithm (LA) in PEDA-Phase-I. By using API pattern recognition, the LA can ascertain whether or not the dubious software was crypto-ransomware. This method may have a high false positive rate (FPR), but it was used to guarantee the most thorough detection of both known and undiscovered crypto-ransomware. PEDA would create a signature of the dubious software and deposit it in the signature repository, which was in Phase II, if the prediction turned out to be crypto-ransomware. By using the signature matching mechanism, the signature repository in PEDA-Phase-II enables the identification of crypto-ransomware at a much earlier level, namely the pre-execution step. Despite being quite strict, this approach was quick and accurate, and it can only identify known crypto-ransomware. To guarantee that

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

the user loses no data, the two PEDA stages created two levels of early detection for crypto-ransomware. But in this study, we concentrated on Phase I, which was the LA. Our findings showed that, when compared to Naive Bayes (NB), Random Forest (RF), Ensemble (NB plus RF), and EldeRan (a machine learning method for analysing and categorising ransomware), the LA had the lowest FPR of 1.56%. A low FPR suggests that LA is unlikely to make an incorrect goodware prediction. To prevent unwanted access to a communication's content, encryption is the act of converting a readable message into an unreadable one. To guarantee that only the intended receiver may access its content, this technique has been frequently utilised in network security. Since the data is readily stolen, it was essential to encrypt it, particularly when it was moving via the network. But it's now established that encryption has drawbacks. On the plus side, it was a great method for cybersecurity to accomplish its goals of non-repudiation, data integrity, data secrecy, and data authenticity. On the other hand, cybercriminals utilised it to demand ransom after locking their victims' data. Crypto-ransomware is an intrusion attack that exemplifies this usage. The term "ransomware" refers to a relatively recent kind of intrusion assault that aims to demand a ransom from its victim. As seen in Figure 1, ransomware mostly comes in three different varieties. Scareware is the first category. The victim of this kind of ransomware is not in any actual danger. Scaring its victim into paying the ransom is its primary goal. Scareware employs a number of tactics, including mimicking an authoritative figure who has discovered the victim's misbehaviour. To avoid being prosecuted, it will seek a ransom. Another somewhat different kind of scareware was dubbed "leakware" because it threatens to reveal the victim's misdeeds to their friends and relatives.

Al-rimy, B. A. S., Urooj, U., and Maarof, M. A. B. (2021, January). A suggested adaptive pre-encryption methodology for early detection of crypto-ransomware. Third International Conference on Cyber Resilience (CRC), 2021 (pp. 1-6). IEEE.

Crypto-ransomware is a kind of malware that encrypts user data using the cryptography features of the system. Unlike other malware types, crypto-ransomware has an irreversible impact, making it difficult to withstand the assault. It becomes challenging to access user data encrypted by a crypto-ransomware assault without the decryption key. Numerous ransomware variations are being created as a result of the availability of ransomware development tool kits such as Ransomware as a Service (RaaS). This helps explain why ransomware assaults are becoming more common these days. However, ransomware cannot be detected using the traditional methods used by malware detection tools. This is due to the fact that ransomware must be identified as soon as possible, even before encryption occurs. Only if these attacks are identified at the preencryption stage can they be successfully countered. The low quantity of data accessible prior to encryption makes it difficult to identify ransomware assaults early. Given the little quantity of data gathered during the pre-encryption stage of the attack lifecycle, this research proposes an adaptive pre-encryption model that should address the population concept drift of crypto-ransomware. With this flexibility, the model can keep abreast of attack patterns and identify polymorphic ransomware, which is malware that exhibits constant behavioural changes. The goal of cybersecurity is to defend networks, devices, and data from online threats. Cyberattacks are initiated with the goal of gaining access to, changing, or destroying user data or intruding into users' business activities. With more devices and fewer personnel, cybersecurity implementation

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

becomes both important and difficult. Attackers are often malevolent individuals that want to access, change, or destroy user information and reputation. Without keeping an eye out for malware assaults, cybersecurity cannot be guaranteed [1]. As intrusion detection systems and anti-virus software have evolved, attackers have also created more potent malware that may adapt and cause serious infections [2]. Malware is a kind of malware created with the intention of harming a computer, server, client, network, or other user's resources. It infringes on user rights and disables or damages computer resources without the user's knowledge. Malware comes in a variety of forms, such as ransomware, worms, Trojan horses, viruses, and rootkits [3]. Attacks using ransomware are a developing trend in the modern day. The extent of the harm it does makes it a severe assault [4]. Ransomware is a kind of sophisticated virus that only has to be run once to get access to your system's resources. It often encrypts all of your files and demands that you accomplish certain tasks in order to unlock the system's resources. It takes use of user permissions. Typically, a ransom demand is made in exchange for cash [5]. The main danger associated with these assaults is that the perpetrator may not always fulfil his commitments and fail to provide the decryption key after payment. In addition to financial loss, the inventor may suffer additional losses including data loss, reputational damage, or even death. Ransomware serves as an incentive to make money and draws malicious minds. Ransomware may also carry out a variety of additional tasks, such killing or stopping programs. In order to prevent the user from working, it may also create a virtual desktop [6, 7]. Money incentive led to an increase in these assaults. Attack-initiating ransomware files are often distributed via Drive-by download, phishing emails, intentional web breach, and weaknesses in systems that are accessible via the internet [6]. The ransomware changed over time. Every day, new variations are being created. Along with various ransomware variations, several ransomware families are created. Several obfuscation techniques, including packing, polymorphism, metamorphism, variable renaming, and trash code insertion, are used to create new variations. The need for adaptive and pre-encryption crypto-ransomware detection systems was brought to light by evolution, growth pace, and sheer volume [8]. Money and simple-to-use kits are the primary factors that fuel the growth of crypto-ransomware. Ransomware as a Service (RaaS) enables even inexperienced attackers to launch a powerful cryptoransomware assault. Existing solutions do not take into account the dynamic nature of ransomware assaults, despite their attempts to identify ransomware early in the pre-encryption stage. The emergence of zero-day attacks complicates detection efforts [9]. In order to identify crypto-ransomware assaults before they begin extortion, an adaptive pre-encryption detection system is necessary [10]. This research aims to solve the deficiency of adaptive preencryption ransomware detection.

II. SYSTEM ANALYSIS

2.1. EXISTING SYSTEM:

A complex and always changing threat, ransomware may encrypt data or lock users out of their computers and demand a fee to unlock them. Crypto ransomware, which encrypts users' data, and locker ransomware, which stops users from accessing their machines, are the two primary categories of ransomware. Conventional methods of ransomware detection, including data-centric, statistical, and event-based methods, are not always successful. Consequently, it is critical that the scientific community create fresh and creative strategies to counter ransomware.

2.1.1 DISADVANTAGES OF EXISTING SYSTEM:

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

- Conventional methods for detecting ransomware don't always work. The scientific community is always creating fresh and creative ways to stop ransomware. Here are a few more drawbacks that may be mentioned:
- For victims, ransomware may be quite expensive. Victims may have to pay for data recovery, IT consultancy, and legal expenses in addition to the ransom.

2.2. PROPOSED SYSTEM:

Malware that encrypts a victim's data and requests a ransom to unlock them is known as ransomware. This study suggests a feature selection-based framework that employs many machine learning algorithms to categorise the security level for ransomware detection and prevention, since machine learning has been shown to be successful in ransomware detection. The suggested methodology uses neural network-based architectures and conventional machine learning classifiers to choose a number of characteristics for model construction. A ransomware dataset is used to test the framework, and the findings indicate that random forest classifiers perform better than other techniques in terms of accuracy, F-beta, and precision scores. The results imply that frameworks based on machine learning may be useful for ransomware detection. Furthermore, the results imply that random forest classifiers could be a particularly successful method for ransomware identification. These results may aid direct the creation of fresh ransomware detection tools that can shield businesses from this escalating danger.

2.2.1 ADVANTAGES OF PROPOSED SYSTEM:

- The foundation of it is transformers, a potent machine learning model that has shown efficacy in sequence-to-sequence tasks.
- The model can learn long-range relationships in the data thanks to self-attention methods.
- It is interpretable, meaning that the characteristics that the model considers may be used to explain the predictions made by the model.

2.3. IMPLEMENTATION:

2.3.1. MODULES:

- User
- Admin
- Data Preprocessing
- Machine Learning

2.3.2. MODULES DESCRIPTION:

User:

The first person to register is the user. For future correspondence, he needed a working user email address and cellphone number while enrolling. The administrator may activate the user once they have registered. The user may log in to our system when the administrator has activated them. Here, we used a dataset that had 138,047 samples with 54 characteristics in total. Of them, 70% were malware, while the remaining 30% were valid observations. The fresh data for the dataset based on our Django application was obtained using the machine learning approach. The user may start the data cleaning process by clicking on the Data Preparations link on the webpage. The information will be shown together with the necessary values.

Admin:

Admin may use his login credentials to log in. The registered users may be activated by the admin. Only the user can log in to our system when he has activated. The administrator may see all of the data in the browser. After the algorithm has finished running, the administrator may see the total accuracy on the website.

Data Preprocessing:

A collection of data items, also known as records, points, vectors, patterns, occurrences, instances, samples, observations, or entities, may

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

be thought of as a dataset. Numerous features that capture an entity's fundamental properties, like the mass of a physical object or the time an event happened, are used to define data objects. Features are often referred to as fields, attributes, dimensions, variables, or characteristics. This forecast's data pretreatment employs methods such as eliminating noise from the data, eliminating missing information, changing default values where appropriate, and classifying features for prediction at different levels.

Machine learning:

On a chosen set of characteristics for ransomware classification, we used a variety of machine learning methods, including Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), and Neural Network (NN)-based classifiers. To test our suggested approach, we conducted all of the tests on a single ransomware dataset. The experimental findings show that RF classifiers perform better than other techniques in terms of precision, accuracy, and F-beta scores.

III. METHODOLOGY

3.1 MACHINE LEARNING ALGORITHMS

Computational models known as machine learning algorithms allow machines—especially computers—to learn from data and make judgements or predictions without explicit programming. These algorithms make use of statistical methods to identify trends, gain knowledge from past experiences, and gradually enhance their performance as they are exposed to further data.

3.2. RANDOM FOREST

An ensemble learning system called Random Forest mixes many decision trees to increase the resilience and accuracy of predictions. The final result is obtained by aggregating the outputs of each tree in the forest, either by average for regression or by majority voting for classification, after each tree has been trained on a random sample of the data and features. Random Forest works well for both classification and regression applications because of its ability to minimise overfitting. It can evaluate the value of features, is flexible, and manages missing data effectively, but it may need a lot of processing power for big datasets.

3.2.1 DECISION TREE

A decision tree is a supervised learning technique that divides data into subsets according to feature values in order to produce a decision tree-like model. It is used for classification and regression problems. A feature test is represented by each node, potential outcomes are represented by branches, and predictions are represented by leaf nodes. Recursively splitting the dataset to maximise class separation or minimise regression error allows the tree to develop. Although decision trees are simple to comprehend, analyse, and display, if they are not regularised or pruned, they may overfit on complicated datasets.

3.2.2 NAÏVE BAYES

Based on the Bayes theorem and supposing high feature independence, Naïve Bayes is a simple but effective probabilistic machine learning technique. In spite of this "naïve" presumption, it works effectively in a variety of applications, including sentiment analysis, spam detection, and text categorisation. The method assigns the class with the greatest probability after calculating the likelihood of a class given the characteristics. Although it is quite effective, handles categorical data well, and performs well with small datasets, its accuracy may decrease if the independence assumption is broken if the characteristics are heavily linked.

3.2.3 LOGISTIC REGRESSION

Binary classification problems are the main application for the supervised learning method known as logistic regression. It uses the logistic (sigmoid) function to estimate probabilities in order to represent the connection between a collection of independent variables and a binary dependent variable. In order to forecast class labels (such as 0 or 1), the algorithm produces

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

probabilities that are then thresholded. For linearly separable data, it is straightforward, interpretable, and efficient. With methods like one-vs-rest, it may be extended to multiclass classification. But until characteristics are changed or paired with non-linear techniques, it has trouble understanding intricate, non-linear interactions.

3.2.4. NEURAL NETWORK

Inspired by the structure and operation of the human brain, a neural network is a machine learning model that can identify patterns and provide predictions. It is made up of layers of linked nodes, or neurones, each of which uses a non-linear activation function and weighted connections to analyse incoming data. Neural networks are appropriate for applications like picture identification, natural language processing, and time-series prediction because they can learn intricate, non-linear correlations in data. Although they are very adaptable and scalable, they need a lot of data and processing power to train, and if they are not adequately regularised, they may be prone to overfitting.

3.3 MATPLOTLIB

A Python 2D plotting package called Matplotlib generates figures of publishing quality in a range of hardcopy formats and interactive environments for various platforms. Matplotlib is compatible with online application servers, the Jupyter notebook, the Python and IPython shells, four graphical user interface toolkits, and Python scripts. It offers several tools for data visualisation and the creation of static, animated, and interactive graphs. For activities ranging from basic line plots to intricate visualisations, Matplotlib is often used.

The Visualization Design using matplotlib

- Bar Graph
- Pie Chart
- Box Plot
- Histogram
- Line Chart and Subplots
- Scatter Plot

3.4 BAR GRAPH

When comparing the number of categorical values inside a single category, bar graphs work well. Continuous numbers shouldn't be shown on bar graphs.

The matplotlib plt.bar() function is used to create bar graphs.

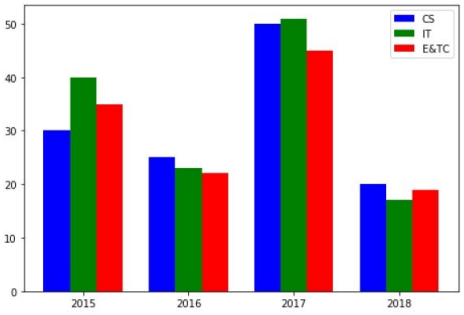


Figure 3.3. Bar Graph

3.5 PIE CHART USING MATPLOTLIB

To display the proportionate distribution of objects within a single category, a pie chart works well.The pie chart is created using plt.pie(), and its settings are changed to make it more visually attractive.

When there are several items in a category, a pie chart becomes meaningless. Each slice will be smaller as a result, and the things won't be able to be distinguished from one another.

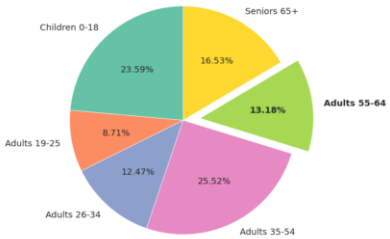


Figure 3.4. Pie Chart

3.6 BOX PLOT USING MATPLOTLIB

A box plot provides statistical details on how numerical data is distributed across several groupings. Finding outliers within each group is one of its uses. The 25th, 50th, and 75th percentile values are represented by the bottom, middle, and top halves of the box, respectively.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746
Vol. 21, Issue 2, 2025

The distribution of data points within each group is not shown using a box plot.

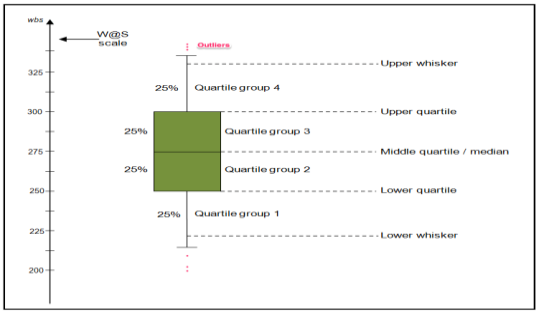


Figure 3.5. Box Plot

3.7 HISTOGRAM USING MATPLOTLIB

By dividing data into distinct bins, a histogram displays the distribution of numerical values across a continuous interval. helpful for examining data skewness. Bar graphs and histograms are often confused. However, keep in mind that bar graphs are used with categorical data, whereas histograms are utilised with continuous data.

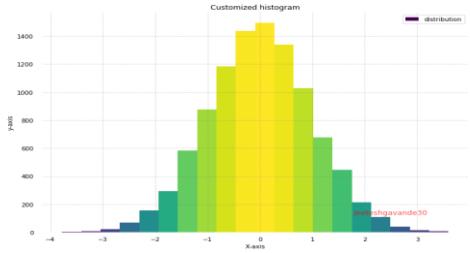


Figure 3.6. Histogram

3.8 LINE PLOT AND SUBPLOTS USING MATPLOTLIB

When displaying the trend of a numerical number over an extended period of time, a line plot is helpful. Plots in the same figure may be easily seen and compared thanks to Matplotlib subplots. The figure and axes are returned by the plt.subplots() function. You may specify how you would want the axes in the figure to be shown as an input to the function. The parameters for nrows and ncols will be used to modify these. The figsize option even allows you to change the figure's size.

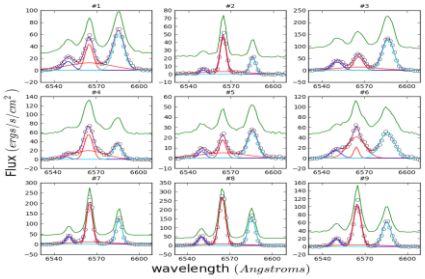


Figure 3.7. Subplot

3.9 SCATTER PLOT USING MATPLOTLIB

When demonstrating the link between two variables, scatter plots are helpful. Scatter plots make it simple to identify any outliers in the data or correlations between variables.

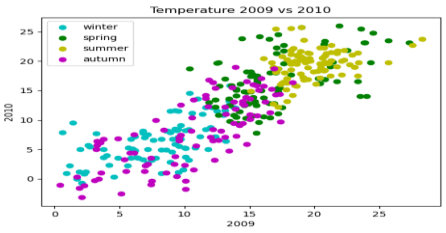


Figure 3.8. Scatter Plot

3.10 SEABORN

Seaborn is a Matplotlib-based toolkit for visualising statistical data. Seaborn is a robust and adaptable Python data visualisation package that provides an intuitive user interface for producing educational and visually appealing statistics visualisations. It offers a variety of data visualisation capabilities, such as sophisticated statistical analysis, and facilitates the creation of intricate multi-plot visualisations. The main advantage of Seaborn is its ability to produce visually appealing plots with little coding work. You may quickly alter its selection of pre-installed themes and colour schemes to fit your tastes. A variety of integrated statistical capabilities are also provided by Seaborn, enabling users to quickly and simply conduct intricate statistical analysis using their visualisations. Seaborn's capability to produce intricate multi-plot visualisations is another noteworthy feature. Users may easily compare different variables or subsets of data by

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746

creating grids of charts using Seaborn. For exploratory data analysis and display, this makes it the perfect tool.

3.11 PLOT TYPES IN SEABORN

- LINE PLOT: Trends in data over time or other continuous variables may be seen using line plots. Each data point in a line plot is joined by a line to form a smooth curve. The lineplot() method in Seaborn may be used to make line plots.
- HISTOGRAM: A histogram shows how a continuous variable is distributed. Data is grouped into bins in a histogram, and the height of each bin indicates the frequency or number of data points in that bin. The histplot() method in Seaborn may be used to construct histograms.
- BOX PLOT: One kind of visualisation that displays a dataset's distribution is a box plot. They are often used to compare how one or more variables are distributed across several groups.
- The violin plot is a kind of data visualisation that blends elements of density plots and box charts. In a box plot-like format, it shows the median, interquartile range (IQR), and density estimate of the data, which is often smoothed using a kernel density estimator. The IQR and median are shown as a white dot and line within the violin, while the breadth of the instrument indicates the density estimate, with bigger sections indicating greater density.

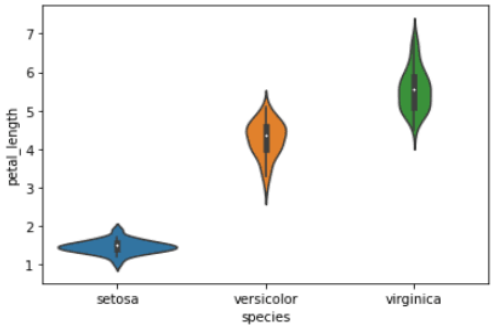


Figure 3.9. Violin Plot

- **HEATMAP**. An illustration of data that employs colours to show a variable's value in two dimensions is called a heatmap. Heatmaps are often used to show how various factors in a dataset are correlated.

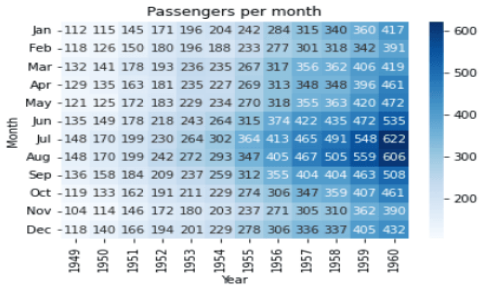


Figure 3.10. Heat Map

- **PAIRPLOT**: Multiple pairwise scatter plots are shown in a matrix style in pair plots, a kind of visualisation. While the diagonal plots display the distribution of the individual variables, each scatter plot displays the connection between two variables.

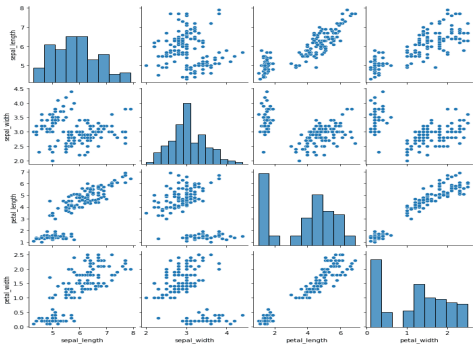
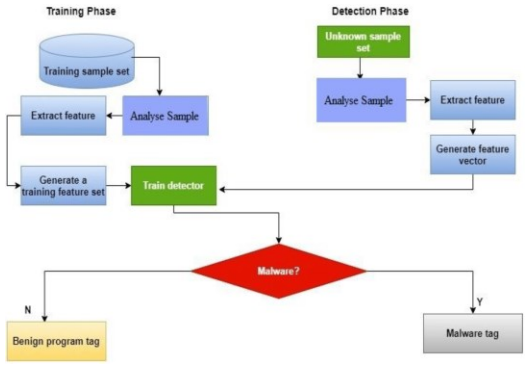


Figure 3.11. Pair Plot

SYSTEM ARCHITECTURE:

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025



IV. CONCLUSION

The study's findings highlight how crucial machine learning is to improving ransomware detection systems. Support Vector Machines (SVM) emerged as the most effective classifier with a high accuracy rate after a thorough evaluation of many techniques. The findings highlight how crucial feature engineering is to improving the model's discriminative skills, especially when it comes to API call attributes. By exploring machine learning, this work adds to the continuous attempts to strengthen cybersecurity defences. Our study intends to provide concrete advantages in the form of more resilient and proactive cybersecurity solutions by concentrating on enhancing the precision and versatility of ransomware detection systems. In addition to assessing how well different machine learning techniques identify ransomware trends, the thorough examination of these techniques provides insightful information that may direct the creation of robust defences against the ever-changing risks presented by ransomware assaults. Our research equips organisations with proactive defence measures and clever algorithms to keep ahead of cyber attackers in their never-ending game of cat and mouse. The collection of algorithms offered in this paper provides a path for the creation of complex and robust cybersecurity solutions, laying the foundation for future developments in the area of ransomware detection.

REFERENCES

[1] Kok, S. H., Abdullah, A., & Jhanjhi, N. Z. (2022). Early detection of crypto-ransomware using pre-encryption detection algorithm. Journal of King Saud University-Computer and Information Sciences, 34(5), 1984-1999.

[2] Kok, S. H., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. (2019). Prevention of crypto-ransomware using a pre-encryption detection algorithm. Computers, 8(4), 79.

[3] Urooj, U., Maarof, M. A. B., & Al-rimy, B. A. S. (2021, January). A proposed adaptive pre-encryption crypto-ransomware early detection model. In 2021 3rd International Cyber Resilience Conference (CRC) (pp. 1-6). IEEE.

[4] Al-Rimy, B. A. S., Maarof, M. A., Alazab, M., Alsolami, F., Shaid, S. Z. M., Ghaleb, F. A., ... & Ali, A. M. (2020). A pseudo feedback-based annotated TF-IDF technique for dynamic crypto-ransomware preencryption boundary delineation and features extraction. IEEE Access, 8, 140586-140598.

[5] Alqahtani, A., Gazzan, M., & Sheldon, F. T. (2020, January). A proposed crypto-ransomware early detection (CRED) model using an integrated deep learning and vector space model approach. In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0275-0279). IEEE.

[6] Y Zakaria, W. Z., Abdollah, M. F., Mohd, O., Yassin, S. W. M. S. M., & Ariffin, A. (2022). RENTAKA: A Novel Machine Learning Framework for Crypto-Ransomware Pre-encryption Detection. International Journal of Advanced Computer Science and Applications, 13(5).

[7] Wang, L., He, R., Wang, H., Xia, P., Li, Y., Wu, L., ... & Xu, G. (2020). Beyond the virus: A first look at coronavirus-themed mobile malware. arXiv preprint arXiv:2005.14619.

[8] Morris, J., Lin, D., & Smith, M. (2021). Fight Virus Like a Virus: A New Defense

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp735-746 Vol. 21, Issue 2, 2025

Method Against File-Encrypting Ransomware. arXiv preprint arXiv:2103.11014.
[9] Wang, Z., Liu, C., Cui, X., Yin, J., & Wang, X. (2022). Evilmodel 2.0: bringing neural network models into malware attacks. Computers & Security, 120, 102807.
[10] Chen, X., Hao, Z., Li, L., Cui, L., Zhu, Y., Ding, Z., & Liu, Y. (2022). Cruparamer: Learning on parameter-augmented api sequences for malware detection. IEEE Transactions on Information Forensics and Security, 17, 788-803.