WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Ransomware Detection Using Machine Learning

Ms. Priyanka N. Kokare[1], Sheefa P. Mulani [2], Saniya A. Pathan[3], Nikita A. Kshirsagar[4],Payal M. Kolekar[5]

[1]Asst. Prof., Department of Information Technology Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati
[2,3,4,5]UG students, Department of Information Technology Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati

ABSTRACT –

Recently, Ransomware has grown in popularity. A widespread and powerful threat to consumer cyber security. It becomes more difficult to successfully protect cyber security systems as its capacity to exploit a variety of attack vectors alterations throughout time. Showing result as, the use of machine learning techniques, especially the Random Forest algorithm, has enormous potential for ransomware detection in the academic and professional worlds. As a result, the main goal of this study is to offer results into ransomware detection frameworks, concentrating on the often used Random Forest technique. It is an ensemble learning technique provides a tool for identifying the constantly changing traits of ransomware. This paper not only examines the foundations of Random Forest but also clarifies the particular uses and benefits of it for ransomware detection. By providing a review of multiple ransomware detection frameworks, this study furthers its contribution to the cyber security community. The report provides useful information about the datasets used in ransomware detection research in addition to illuminating the key elements of these frameworks. It also highlights the distinct difficulties each framework can run into when trying to effectively identify the wide range of ransomware. In conclusion, this research presents a comparison investigation, with a particular a focus on the Random Forest algorithm, which serves as a useful resource for future study in identifying ransomware. The benefits and drawbacks of Random Forest are explained in this context, providing academics and experts in cyber security with the knowledge they need to keep on top of the continuous against ransomware attacks.

Keywords: Ransomware, cyber security threat attack vectors, machine learning algorithms, Random Forest.

INTRODUCTION :

1.1. Security :

Security is like a shield that keeps our phones, computers, and online account safe from thieves and harm. It's what helps us uses the internet and our devices without worrying
about someone trying to steal or damage our stuff. Security in a simple term related to viruses, means protecting your computer or device from harmful software that causes damage.
Security is the practice of protecting something valuable from potential threats, dangers, or harm. In various contexts, it can refer to safeguarding information, assets, systems, or even physical spaces from unauthorized access, damage, theft, or any form of malicious intent.

Malware:

Malware is bad software made by cyber hackers it's like a digital troublemaker that can steal your stuff or mess up your computer. Malware is like a digital troublemaker, it short for
―malicious software‖. It's a type of computer program designed by bad actors to do harmful things to your device or steal your information. Malware can sneak into your computer, phone, or other devices without you knowing, and then it can cause problems like slowing things down, stealing your personal info, or even damaging your files. It's like a sneaky virus for your technology! Exactly Malware is like a sneaky program that can sneak into your device and cause all sorts of trouble. It might want to steal your money or personal info, or even do something harmful later on. So, it's best to avoid it altogether!

Types of Malware : -

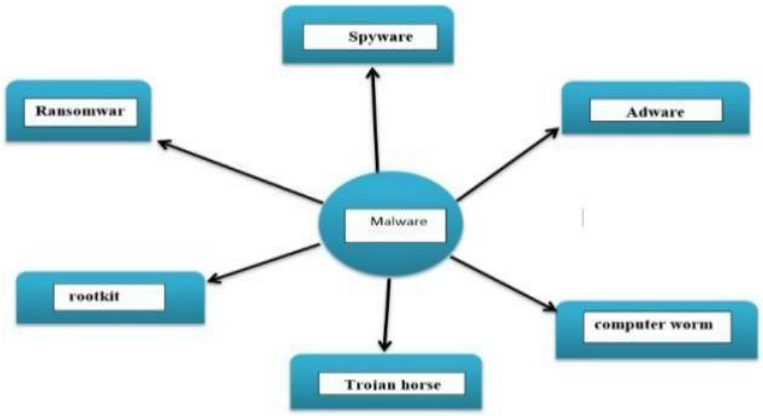

Fig. 1.1

Types of malware:

Ransomware:

Ransomware detection is the attack based on locked encryption key and attackers demands to pay the charges For encryption key. Ransomware is most descriptive cyber threat that causes critical pecuniary losses while impacting productivity, accessibility and relation, ransomware attack is designed by hackers to deny users and organization accesses to files on their computer. Ransomware scan any local device and any network connected storage
In recent years, ransomware has become a menace to government agencies, business units, and regular end consumers. As instances, it has attacked hospitals[1], universities[2], school districts[3], police departments, and school departments.

There are two categories into which ransomware falls:

1 **Locker -ransomware-** It is suggests that the victims' gadgets be locked to keep them away from using them.
2 **Crypto -ransomware-** In order to prevent their victims from accessing personal files, this encrypts them.

Spyware:

Cyber security on an individual and organizational level is seriously threatened by spyware, which infiltrates computers covertly to steal confidential data. This article will examine the top five most well-known spyware assaults in history, including illuminating data and research to comprehend their significance and aftermath.
Eg. Finspy(Finfisher)

Adware:

Malicious software known as "adware" breaks into your system and shows pop-ups and unsolicited adverts. Adware is able to track what you do online and present you with tailored adverts. The word "adware" refers to software that, typically via a web browser, displays advertising on your screen. Some security experts believe it to be the ancestor of the modern PUP (potentially unwanted software). Malware typically uses deception to trick you to install program on your device. It may do this by seeming to be a reliable piece of software or by taking advantage of another one.

Computer Worm:

The type of malicious software that propagates to other systems is called a computer worm. By self- replicatingand frequently damaging data or taking up bandwidth. Eg. Morris Worm

Trojan Horse:

This code has the ability to take over the computer and is malicious in nature. It is intended to cause harm, theft, or other undesirable activities on the computerIt tries to fool the user into letting the files load and operate on the device. This provides hackers access to the user's computer once it is operational to carry out a variety of tasks, including deleting and editing data from folders. Trojan Horse, like many other viruses or worms, is unable to replicate itself.

Rootkit:

Software of the rootkit malware class is intended to grant unauthorized users access to a computer network or program. After activation, the malicious program installs a backdoor exploit and can propagate various viruses, including ransomware, trojans, bots, and keyloggers. Rootkits can be hard to find and remain unnoticed for years because they can interfere with various antivirus and malware detection software.
E.g. Firmware rootkit

History of Ransomware:

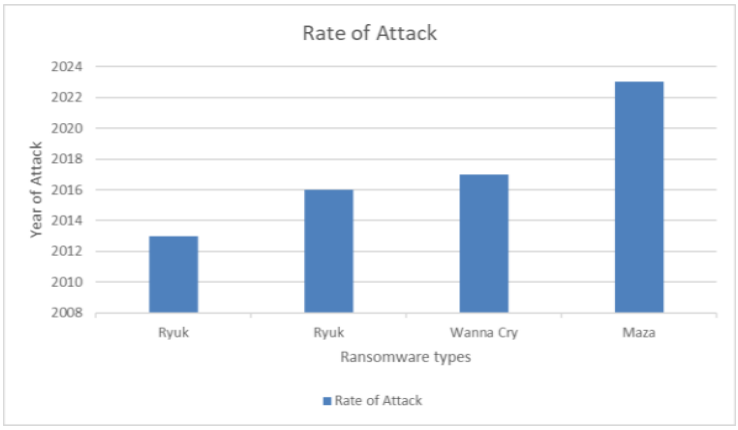

Fig.1.2. History of Ransomware

As shown in above graph ransomware has evolved day by day and has been very danger in malware. Hence, ransomware needs to be detected before it harms whole system. Every day, the frequency of ransomware attacks rises that needs to be controlled.
Recent ransomware attack have corrupted many computers. Ryuk, Wanna Cry and Meta are ransomwares which has become harmful.

1. **1989(AIDS Trojan):**
- It was developed by Dr. Joseph_Popp and spread over 20,000 participants of the WHO AIDS seminar.
- $189 ransom was demanded.

2. **2004/2005(GPCoder):**
- A notification that points users to a text file (.txt) that is put on their desktop appears on their home screen. The file included instructions on how to unlock the compromised file and pay the ransom.
- •Demanded a ransom of $200.

3. **2006(Archiveus Trojan):**
- Primarily an attack that runs on Windows.

4. **2009(Locker ransomware):**
- A type of ransomware that targeted mobile gadgets.
- Notable instances are Reveton and WinLock.

5. **2013(CryptoLocker):**
- The first ransomware to demand bitcoin as payment.

6. **2014(Cryptowall):**
- Made use of a Java exposure; around 1,000 fatality; minimum drop predicted to be $18 million.

7. **2016(Locky):**
- The 1st ransomware went viral; up to 500,000 phishing emails were sent out every day.

8. **2016 also saw the introduction of more ransomware, including as Cerber, Jigsaw, SamSam and Petya.**

9. **2017(WannaCry and NotPetya):**
- Approximately 200,000 systems across 15 countries were targeted by WannaCry. Notable victims include WinLock and Reveton.
- A Petya variation known as NotPetya was directed towards targets in Ukraine, such as the National Bank of Ukrain.

10. **2021(DarkSide):**
- Following an attack that shut down the pipeline for 6 days, Colonial Pipeline rewarded a $4.4 million bitcoin.

An analysis of the most pertinent findings about ransomware detection:

Table provides a comparison of various research results related to the detection of ransomware. The table shows several authors' work over different years and their use of various techniques for ransomware detection. KNN, SVM, Decision tree, AES, Local n/w, Dynamic n/w, CNN, RNN, RF, k-means, Crypto-currency algorithms, NB, LSTM, Linear model: These columns represent different techniques or algorithms used for ransomware detection. A checkmark (✓) indicates that the technique was applied, while a cross (×) indicates that it was not used in that particular research.
These are the techniques and methods applied by different researchers in their respective years to address the challenge of ransomware detection. Each approach may have its own strengths and limitations, and the table provides a quick overview of the methods used in each study.

Sr. No	Year	Author	KNN	SVM	Decision tree	AES	Local n/w	Dynamic n/w	CNN	RNN	RF	k-means	Crypto-currience algorithms	NB	LSTM	Linear model
1	2019 [14]	K.Lee	✓	×	✓	×	×	×	×	×	×	×	×	×	×	✓
2	2022 [15]	Faimah aldauui	×	×	×		×	×	×	×	×	×	×	×	×	×
3	2019 [16]	Edurdo Berrueta	×	×	×	×	✓	✓	×	×	×	×	×	×	×	×
4	2020 [17]	Emad Badawi	×	×	×	×	×	×	×	×	×	×	✓	×	×	×
5	2023 [18]	Salwa Razautla	×	×	×	×	×	×	✓	✓	✓	×	×	×	×	×
6	2019 [19]	Anurag zen	×	×	×	×	✓	✓	×	×	×	×	×	×	×	×
7	2020 [20]	Trailor	✓	✓	×	×	×	×	×	×	×	×	×	×	×	×
8	2019 [21]	Tandom	×	×	×	×	✓	✓	×	×	×	×	×	×	×	×

Machine Learning:

Machine learning is derived from AI. It pivots on statistical models and strategy that allows computers to get knowledge from data and improve their effectiveness on certain work unaccompanied by the need for explicit programming.

1. Learning from Data: Machine learning algorithms use vast datasets to discover patterns and relationships that enable computers to anticipate outcomes, make choices, or spot patterns in previously undiscovered data.

Training and Testing: 1. Machine learning models are trained on labeled datasets, which educate the computer how to convert input data into matching output labels.

2. To determine the model's accuracy and generalization capacity, its performance is assessed on testing data, which is a different set of data not used for training.

3. Types of Learning: Machine learning includes reinforcement learning, which teaches algorithms to make decisions by trial and error, supervised learning, which teaches algorithms from labeled data, and unsupervised learning, which teaches algorithms to find patterns in unlabeled data.

4. Methods and Models: Support vector machines, decision trees, and neural networks are just rare of the methods used in ML. Models are provided by these algorithms with the ability to classify or predict things based on brand-new, unobserved data. Machine learning is a powerful instrument for managing complex problems and creating data-driven judgments in a variety of fields since it allows computers to learn from experience.

1.7 Random Forest: Typically trained with the bagging method. An technique Random Forest, a method for supervised learning, builds a forest of decision trees. It often has the same hyperparameters as a decision tree and can be applied for regression and categorization tasks. In an attempt to produce an even better model, this method generates trees while increasing the model's randomness. When dividing a node, it seeks for the top feature from a arbitrarily selected set of features instead of the most notable one. Using random thresholds for each attribute
instead of looking for the optimal thresholds can make trees appear more haphazard. When developing a model, random forest is an excellent approach to train and observe its performance. Its simplicity makes creating a decent random forest quite easy.

Decision Tree : Tasks involving regression and categorization can be done with decision trees. Here with Ransomware detection in Machine learning can work. This system doesn't prove much better because it is prone to be overfitting. When it is trained with with more data and then tested it cannot accurately recognize ransomware. Recognition of variety of ransomware families with use of decision tree algorithm proves to be difficult. Hence, it is not much chosen by researchers.

1.10. Support Vector Machine: Although it may be used for both Regression and classification issues, the Support Vector Machine approach is mainly employed for categorization goals. It generates results with amazing accuracy while requiring relatively little computing resources. In N-dimensional space, the support vector machine approach locates a hyperplane that classifies the data points. The hyper-plane that most successfully separates the two classes is identified throughout the
categorization procedure. [26]. Nevertheless, because they perform well in high dimensional regions, they can be useful in circumstances when there are more dimensions than samples. Another important feature of the method is its adaptability; it can employ bespoke kernels or a variety of kernel functions against the decision function. On the other hand, overfitting will happen if there are significantly more characteristics than samples.

LITERATURE SURVEY

Paper No.	Paper Name	Author Name	Proposed System	Algorithm Used	Pros	Cons	Accuracy
[1]	RTrap: Trapping & containing ransomware with ML.	Gaddis Olani Ganfure, Yan-Hao Chang et al[1]	Ransomware detection using adaptive decoy file generator	K-means	Easy trapping of ransomware	Unable to detect	96.5%
[2]	On ransomware family attribution pre-attack activities	Sadegh Tarabi, Ayala Molina et al[2]	Ransomware samples in collected data are labeled.	Bernouli Naïve Bayes, KNN	Early detection of ransomware attack	Doesn't handle sustainability issue.	78%
[3]	Machine Learning Algorithm and Framework in ransomware Detection	Daryle Smith, Sajad Khorsandroo, Kaushik Roy et al[3]	Tested against several framework and ransomware types.	Decision Tree, Random Forest, SVM, Naïve Bayes, LSTM, Gradient Boosting.	1. Enhanced detection 2. Automate detection	Requires large amount of data.	94%
[4]	Ransomware detection using Random Forest Technique	Ban Mohammad Khammas et al[4]	offers a unique architecture for resolving the ransomware attack detection challenge.	Random Forest	1. It has high accuracy rate & less time consuming	Imbalanced dataset	97.74
[5]	A Comprehensive Survey: Ransomware Attacks Prevention, Monitoring and Damage Control	Ashish D. Patel et al[5]	Provides detection and prevention of ransomware	Tools like CryptoDrop offer early detection of suspicious activities.	Easy trap.	Requires large data.	N.A.
[6]	A Survey on Detection Techniques for Cryptographic Ransomware	Daniel Morato et al[6]	Algorithms that have been suggested for identifying cryptographic ransomware.	Compared the different approaches and classified the algorithms.	Good performance.	Discussed the open issues to offer solutions for ransomware detection.	96%
[7]	Detection of ransomware attacks using processor & disk usage	Mehnaz et al [7]	A detection method for hybrid ransomware was presented by the author.	SVM	Tracks malware activity on the intended machine directly.	Authors did not explicitly investigate data by ransomware	96.22%

[8]	On ransomware family attribution using pre-attack paranoia.	Zhang et al, Subedi et al [8]	Data collection, preprocessing, and labelling of malware samples.	Decision tree, RF, SVM, ANN, CNN.	Allows ant i- ransomware system	System requires large	88.76%
[9]	A Digital DNA Sequencing Engine F or ransomware detection using ML.	Kharraz , Shaukat an d Ribeiro et al[9]	Proposed RansomWal l	Logical regression classifier algorithm			82%
[10]	Ransomware detection, avoidance, and	Subedi , Zheng ,Zimba et al[10]	Approach that makes use of static analysis is suggested.	Algorithm for Pre-Encryption Detection	All crypto ransomware might be identified by PEDA.	It could be unable to identify the	78.25%
	mitigation scheme: A review an d future directions			(PEDA)		newest families.	

Paper No.	Paper Name	Author Name	Proposed System	Algorithm Used	Pros	Cons	Accuracy
[11]	Learning to detect and classify malicious executables in the wild	LO et al, Schultz et al, Krsul, Swets and Pickett et al[11]	They collected safe and harmful programs to analyse	Boosted decision trees, SVM, Naïve Bayes.	They attempted to extract patterns or signatures to identify any class of malicious code		89%
[12]	Ransomware detection using Random Forest Technique (Machine Learning)	Abhijit Pawar, Pradnya Kapse, Rutuja Jagtap, Pooja Onigude, Rupali Deorate et al[12]	Collect diverse data, engineer feature importance, optimized deploy	Random Forest, Supervised learning algorithm, semi-supervised learning algorithm and SVM	Few input parameters required	The resulting ensembles model can be complex	96.75%
[13]	Malware Detection: A framework for reverse engineered android application through Machine Learning	Beeish Urooj, Munam Ali Shah, Riasat et al[13]	It contains 2 parts: 1. Preprocessing 2. Prepared Model	Support Vector Machine, Decision Tree, KNN, Naïve Bayes	1. Automation and speed for processing many apps 2. Scalability to handle large dataset	Dependence on high quality and diverse training data	96.24%
[14]	Machine learning based file entropy analysis for ransomware detection backup system	K.LEE, S.lee k.yim et al[14]	The backup system recognizes files contaminated by reference values using a machine learning model.	KNN, Linear model, Decision tree, neural network, entropy	Quickly identifies ransomware pattern	For effective training, a significant numbered label information may be needed, that	97%

						might be hard to come by, particularly for uncommon or newly appearing ransomware variants.	
[15]	Utilizing cyber threat trunting techniques to find ransomware attack and survey of art	FAIMAH ALDAUUI, OMAR BATARFI MANAL BAYOUS, MAVROEIDS et al[15]	Provide an automatic threat assessment system that categorizes the system by analyzing the constant incoming feeds from the Sysmos logs.	AES encryption algorithm	Accuracy issue.	The condition redundancy phrase is not taken into account in the proposed work when determining the feature important	68%
[16]	A survey detection technique for cryptographic ransomware	EDUARDO BERRUETA, DANIEL MORA TO MIKEL IZAL et al[16]	This method for detecting ransomware has been suggested by the academic and industrial sectors.	Local static, local dynamic network	The method for detecting ransomware that is aware of its behavior during one or more of the processes listed above	Common key for any interface user Symmetric key algorithm is the same as encryption key	N.A.

Paper No.	Paper Name	Author Name	Proposed System	Algorithm Used	Pros	Cons	Accuracy
[17]	cryptocurrencies Emerging Threat and Defensive mechanism: A systematic literature Review	EMAD BADAW, W ed the GUY- VINCENT et al [11 7]	We used the standard guidelines and charts for systematic literature reviews in this investigation.	Crypto currencies clustering algorithm	Simple detection.	Require s time.	80.76%

[18]	A age of ransomware a survey on the evolution taxonomy and research direction	SALWA RAZAUL LA, CLAUDE FACHKHA, et al [18]	In order to arrive at a categorization and conclusion, Using characteristics extracted from ransomware activity, algorithms were compared and categorized.	CNN, RNN, SVM	Ability of ML, which includes DL, can recognize trends that indicate the presence of ransomware by learning from historical data.	Slows down.	98.45%
[19]	Ransomware detection avoidance and mitigation scheme: a review and future direction	Anurag zeb et al [19]	Analyze attack strategies for families of Windows-based ransomware.	Static, dynamic, hybrid analysis	The author covered every potential vulnerability vector and kit that might be used to create a family of ransomware that targets Windows.	They made no mention of the technical solutions needed to combat ransomware.	N.A.
[20]	A comprehensive survey on ransomware attack: prevention monitoring and damage control	Trailor et al [20]	Examining various encryption methods employed by contemporary ransomware standards in order to create more effective detection techniques.	K-MEANS, KNN. SVM	The author provided a thorough analysis of the various encryption methods employed by the crypto ransomware and locker families.	Talks are given in relation to a single ransomware and broad-based containment techniques. Not recommended.	94%

Paper No.	Paper Name	Author Name	Proposed System	Algorithm Used	Pros	cons	Accuracy
[21]	A comprehensive Survey ransomware attack: A growing havoc cyber threat	Tandon et	described the typical ransomware attack's architecture and mode of operation.	Local, dynamic network	The author provides a detailed account of the MS017 exploit and how it ultimately employed double dolman to propagate Wannacry.	The discussion is limited to a single ransomware board and does not include any countermeasure techniques.	NA
[22]	Know Abnormal, Find Evil: Frequent Pattern Mining for Ransomware Threat Hunting and Intelligence	Sajad Homayon, Ali Dehghantaha, Marzeih Ahmadzedh, Sattar Hashemi et al[22]	Technique for Mining Sequential Patterns used to identify the most useful characteristics for categorizing ransomware applications.	Random Forest, J48, Bagging and MLP algorithm	Gave much accurate result to goodware and ransomware.	Slow for detection of ransomware	99%

[23]	Automated Analysis Approach for detection of high survivable ransomware	Yachya Abukar Ahmed, Barus Kocer, Bander Ali Saleh Al-rimg et al[23]	Contains 3 stages: 1. Checking whether ransomware or not. 2. Analyze samples Supervised learning algorithm for classification	SVM and ANN to develop and implement decision model. k-nearest neighbour, decision tree and random forest for evaluation.	Provides high accuracy.	Sometimes for different types of ransomware gives less accuracy.	98.7%
[24]	RATAFIA: Ransomware Analysis using Time and Frequency Informed Autoencoder	Manner Alam, Sarani Bhattacharya, Swastika Dutta, Sayan Sinha, Deebdeep Mukhopandhyay and Anupam Chattopadhyay et al[24]	RATAFIA, which develops highly accurate, quick, and dependable ransomware detection solutions, makes use of deep neural networks and fast Fourier transformation.	Deep Neural Network	A concept to recover ransomware-encrypted files before detection by using the Linux file locking mechanism, the mlock () system call.	An adversary cannot alter the execution footprint of its own page or evade detection schema if it is aware of the template used to train the autoencoder.	96.27%
[25]	RANDS: A Machine Learning-Based Anti-Ransomware Tool for windows platform	Hiba Zuhair and Al Selamat et al[25]	3 tier technology: 1. Ransomware Analysis 2. Learning tier 3. Detection tier	Naïve Bayes and Decision Tree	In a brief amount of time and with little impact on the computer system, RANDS addressed ransomware variants in terms of CPU usage during processing, detection, and reaction times.	1. Does not adequately describe all families and exploitations of ransomware. 2. A variety of detection strategies that guaranteed distinct deciding principles for recognizing ransomware variations.	96%

Accuracy analysis from table1:

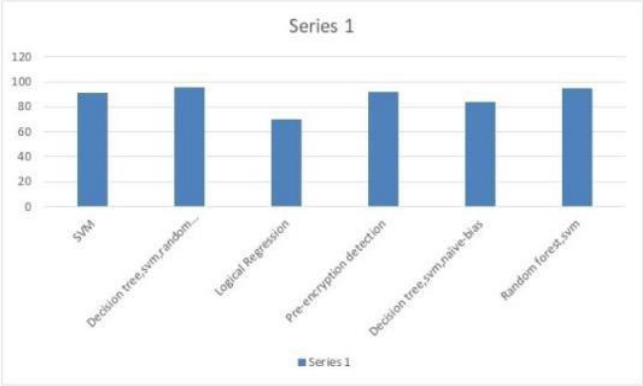

Fig.1.3. Accuracy analysis from table1

Accuracy analysis from table2:

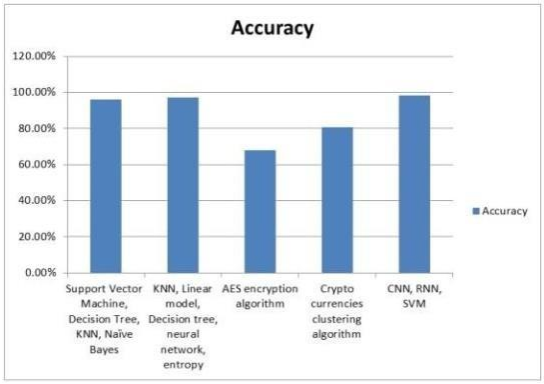

Fig.1.4. Accuracy analysis from table2

Accuracy analysis from table3:

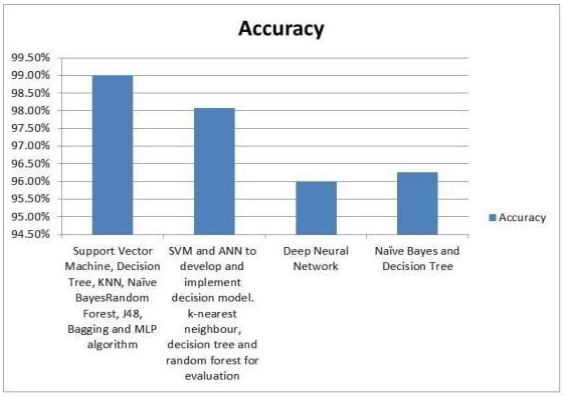

Fig.1.5. Accuracy analysis from table 3

Proposed Work

a) **System Architecture:**

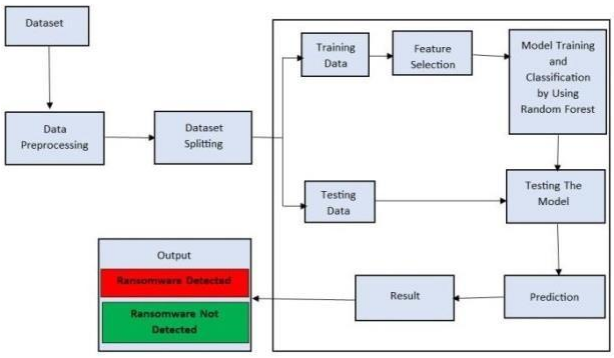

Fig.1.6. Proposed System Architecture

b) **System Design:**

1. **User uploads file**: The process begins with uploading a file into the system. This serves as the input for the Ransomware detection process.

2. **Data Preprocessing:** To ensure accurate results, the first step in data processing was cleaning the data. This included replacing missing numbers, reducing noise in the data, locating and eliminating outliers, and eliminating discrepancies.
3. **Data Splitting**: Splitting the dataset into two portions will allow you to train the model and measure its effectiveness.

4. **Feature Selection:** The important features (characteristics) in the dataset that will be used to train the model should be identified. This process reduces computational costs while enhancing the accuracy of the model. An ensemble learning technique is the algorithm known as Random Forest. Different decision trees are formed while data being trained process. Random subsets of
5. the characteristics and data are used to train each tree.
6. The model‘s accuracy and overfitting both are enhanced by this diversity.
7. **Model Training and Classification:** Use of a Random Forest (RF). The Random Forest method may identify between legitimate and malicious files or network traffic by learning patterns in the data. With features such as major image version, main OS version, major linker version, and minor linker version capabilities playing a crucial part in achieving the best results, RFs are the most effective algorithm for ransomware detection. For each sample in the testing dataset, the model will produce predictions indicating whether it thinks the file or network traffic is ransomware or not. tested dataset was used to gauge the model‘s performance.
8. **Result Generation:** Based on the analysis performed by Random Forest ,the system generates a result that indicates whether the uploaded file is ransomware or goodware file. **Output Display:** The user interface of the system presents the final product to the user.

Algorithm Working: A.Random Forest Working: Random Forest works on ransomware detection by training on a dataset containing examples of ransomware and normal behavior. It creates number of decision trees, each utilizing random subsets of data and features. These trees vote on whether an input is malicious or not. The final decision is based on the most popular prediction. The accuracy and robustness of this ensemble technique are enhanced in detecting ransomware attacks.

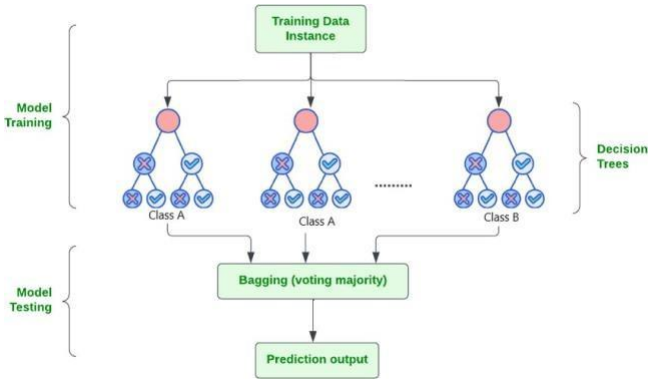

Fig. 1.7. Random Forest

How Does a Random Forest Operate?

The random forest algorithm has multiple stages of operation:

1. **Ensemble of Decision Trees:** It creates a massive army of Decision Trees to optimize the benefits of collaborative learning. Every one of these trees represents a different expert focused on a subset of the data. Their independence is essential because it decreases the possibility that the shades of one tree could negatively impact the representation.

2. **Random Feature Selection**: It selects random features to makes sure that every decision tree in the cluster has a special view. A random portion of the features selected while being trained is available in each tree. Owing to this unpredictability, many predictors are employed, with each tree focusing on a distinct subset of the data.

3. **Bootstrap Aggregating, or Bagging:** An essential step in Random Forest's training process is bagging, which is the process of creating numerous bootstrap specimens from the source dataset so that replacement occurrence can be represented. Because of this, every decision tree contains a unique subset of data, which improves the model's performance and adds variation to the training process.

4. **Decision Making and Voting**: Each Random Forest decision tree gets a vote for a forecast for the future. In categorization tasks, the average, or most frequent forecast, across all trees is used to determine the final prediction. The average forecast for every tree is calculated in regression tasks. The impartiality and cooperation of decision-making are guaranteed by this internal voting process.

B. Tfidf: The abbreviation for term frequency is Tfidf. Record frequency of documents in reverse. this is the procedure for determining appropriate term a text in a dataset. A term significance increases with its number of appearances in the dataset; nevertheless, it is offset by the term occurrence in the dataset.

Term Frequency: This term shows the frequency of a term t, shown in document d, whether frequently or not. Therefore, the most of times a term shown in the text, the most relevant it becomes. Since this ordering is meaningless, we can apply a vector to convey the text in the bag of word models. A document's weight is directly correlated with its frequency of occurrence.

tf(t,d) = count of t in d / number of words in d.

Document Frequency: This is similar to TF in that it determines the text's meaning over the entire corpus. The sole distinction is that in document TF, the term t's occurrence number counter is located, while in document d, the term t's number of appearances in text set N is located in df. Stated differently, the phrase the frequency of t in text, or df(t), appears in certain DF publications.

Inverse Document Frequency: It prioritizes determining relevancy of a word. The primary objective of the exploration is to find the pertinent papers that satisfy the objective. Furthermore, TF does not utilize term frequencies alone to assess a weight of term in the article because it considers every expression to be evenly appropriate. Count the number of texts that contain the expressions to determine its document occurrence.

df(t) = N(t) in which df(t) is the term t's document frequency N(t) = Total number of texts where the phrase "t" appears

C. Support Vector Machine(SVM):): SVM, a supervised ML technique, has two applications: regression and classification. Regression problems work best, nevertheless, when applied to classification problems. Searching the most appropriate hyperplane to partition data points into unique feature space classes in an N- dimensional portion is the basic objective of the SVM technique. The hyperplane's goal is to continue the biggest buffer between the closer points of different classes. The feature number decides the hyperplane's dimension. The hyperplane can be viewed as a line when there are 2 input characteristics. The hyperplane alters into a 2-D plane if there are 3 inputs.

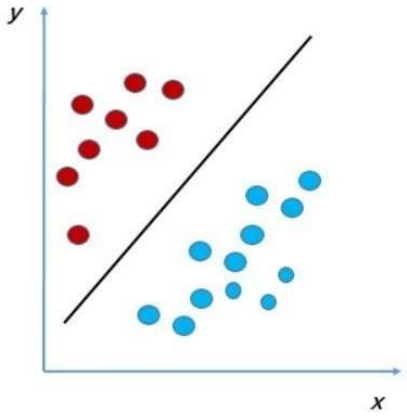

Fig.1.8. Support Vector Machine(SVM)

Challenges

1. Noiseless ransomware cannot be detected.
2. Classifies only familiar Ransomwares.
3. Some evidence that ML algorithms can better identity malware.
4. The performance drops with additional features introduced.
5. Performance can be improved by continuously tweaking the active learning algorithm's parameters.
6. Impacts on file recovery.
7. Windows based platform only.
8. Differentiate between legal and malicious disk encryption procedures.

Performance Measure

Sr no.	Paper title	Formula	Precision	Recall	ROC	TPR	FPR	FNR	F1-score	Time
1.	RTrap:Trapping & containing ransomware with ML.	1. FPR=FP/FP+TN					0.2%			5.35s

2.	Ransomware detection using Random Forest Technique	1. TPR=TP/TP+FN 2. FPR=FP/FP+TN 3. Precision=TP/TP+FP 4. Recall=TP/TP+FN 5. Accuracy=(TP+TN)/(TP+FP +TN+ FN)			99.6%		0.04%	0.002 %		1.37s
3.	Detection of ransomware attacks using processor & disk usage.	1. TPR=TP/TP+FN 2. FPR=FP/FP+TN 3. Precision=TP/TP+FP 4. Recall=TP/TP+FN 5. Accuracy=(TP+TN)/(TP+FP +TN+ FN) 6. F1 score=2TP/2TP+FP+FN	0.970%		0.97%		0.03%	0.030 %	0.960%	400ms
4.	On Ransomware Family Attribution Using Pre-Attack Paranoia Activities	1. Precision=TP/TP+FP 2. Recall=TP/TP+FN 3. F1 score=2TP/2TP+FP+FN	57.69%	54.54 %					94.92%	0.344s -
5.	A Digital DNA Sequencing Engine For ransomware detection using ML.	1. TPR=TP/TP+FN 2. FPR=FP/FP+TN				85%	15%			
6.	Learning to detect and classify malicious executables in the wild	1. FPR=FP/FP+TN					0.05%			
7.	Machine learning-based detection of ransomware using SDN	1. FPR=FP/FP+TN 2. Precision=TP/TP+FP 3. F1 score=2TP/2TP+FP+FN	0.987%				12.5%		0.86	
8.	Know Abnormal, Find Evil: Frequent Pattern Mining for Ransomware Threat Hunting and Intelligence	1. Precision=TP/TP+FP 2. Recall=TP/TP+FN 3. F1 score=2TP/2TP+FP+FN	86%	30%					40%	
9.	Automated Analysis Approach for detection of high survivable ransomware	1. FPR=FP/FP+TN			0.987%		0.007 %			8ms
10.	RATAFIA: Ransomware Analysis using Time and Frequency Informed Autoencoder	1. FPR=FP/FP+TN					0			5s

Sr no.	**Paper title**	**Formula**	**Precision**	**Recall**	**ROC**	**TPR**	**FPR**	**FNR**	**F1-score**	**Time**

11.	RANDS: A Machine Learning-Based Anti-Ransomware Tool for windows platform				0.94					0.2s
12	Malware Detection: A framework for reverse engineered android application through Machine Learning	FPR=FP/FP+TN					0.3			
13.	Machine learning based file entropy analysis for ransomware detection backup system	1. TPR=TP/TP+FN 2. FPR=FP/FP+TN 3. Precision=TP/TP+FP 4. Recall=TP/TP+FN 5. Accuracy=(TP+TN)/(TP+FP +TN+ FN) 6. F1 score=2TP/2TP+FP+FN	high	high	high		low	low		
15.	Utilizing cyber threat trunting techniques to find ransomware attack and survey of art									5s
16.	cryptocurre n- Cies Emerging Threat and Defensive me- Chanism: A systematic literature Review	1. TPR=TP/TP+FN 2. FPR=FP/FP+TN 3. Recall=TP/TP+FN		96.9		95%	4.9%			10s
17.	A age of ransomware a survey on the evolution taxonomy and research direction	1. FPR=FP/FP+TN					0.58%	1.5%		

CONCLUSION

Now a days ransomware has been at high point in cyber security. It is crucial to detect ransomware before it infects your machine. The paper provides information for detecting ransomware. Though various research papers are available many have drawbacks like accuracy. The goal is to detect ransomware before it infects machine with much accuracy. This paper has provided an overview of several different methods proposed by different researchers to addresses the challenge of Ransomware Detection. Many algorithms like SVM, Decision tree are available to detect ransomware. Here, in this paper Random Forest is used as it chooses best decision tree for detection of ransomware.

In conclusion, this review highlights the importance of Ransomware Detection. Detecting ransomware families is somehow challenge and according to many researches system fails here.

REFERENCES :

1. K. Lee, S. Lee, and K. Yim, __Machine learning based file entropy analysis for ransomware detection in backup systems,‘‘ IEEE Access, vol. 7,pp. 110205–110215, 2019.

2. Faimah aldauui, Omar batarfi https://ieeexplore.ieee.org/iel7/6287639/9668973/09791234.p df

3. Eduardo berrueta, mikel izal A Survey on Detection Techniques for Cryptographic Ransomware http://dataset.tlm.unavarra.es/ransomware/articles/IEEEAcces s.pdf
4. Emad badawl, Cryptocurrencies Emerging Threats and Defensive Mechanisms
5. https://ieeexplore.ieee.org/iel7/6287639/8948470/09243940.p df

6. Salwa razauia " a age of ransomware: a survey on the evolution taxonomy and research direction" https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10105 244

7. Anurag zeb Ransomware Detection, Avoidance, and Mitigation Scheme: A Review ... https://www.mdpi.com/2071-1050/14/1/8

8. Trailor,Ransomware Attacks Prevention, Monitoring and Damage Control
9. https://www.researchgate.net/publication/321161261_A_Co mprehensive_Survey_Ransomware_Attacks_Prevention_Mo nitoring_and_Damage_Control
10. tandon, a.; nayyar, a. a comprehensive survey on ransomware attack: a growing havoc cyberthreat. in data management, analytics and innovation; springer: Singapore, 2019; pp. 403–420.

11. Ban Mohammad d khammas ,"Ransomware Detection using random forest techniques" https://www.researchgate.net/publication/346882787_Ransom ware_Detection_using_Random_Forest_Technique

12. Ashish D. Patel A Comprehensive Survey on Ransomware Attack - Springer https://link.springer.com/chapter/10.1007/978-981-13-1274- 8_31

13. Zhang, Subedi On Ransomware Family Attribution Using Pre-Attack Paranoia Activities https://ieeexplore.ieee.org/document/9536608

14. Beeish urooj, munam Ali shan,risat "malware detection a framework for reverse engineered Android application through machine learning"
15. https://ieeexplore.ieee.org/document/970337512)
16. Mehnaz "Detection of ransomware aatack using processor and diskusage"
17. https://ieeexplore.ieee.org/document/10132856/#:~:text=The% 20current%20approaches%20to%20detect,and%20corrupt%20 the%20collected%20data

18. F. Khan, C. Ncube, L. K. Ramasamy, S. Kadry, and Y. Nam,__A digital DNA sequencing engine for ransomware detection using machine learning,‘‘ IEEE Access, vol 8, pp. 119710–119719,2020.

19. J. Z. Kolter and M. A. Maloof, __Learning to detect and classify malicious executables in the wild,‘‘ J. Mach. Learn. Res., vol. 7, pp. 2721–2744,Dec. 2006.

20. Ahmed, B. Kocer, and B. A. S. Al-rimy, __Automated analysis approach for the detection of high survivable ransomware,‘‘ KSII Trans. Internet Inf. Syst., vol. 14, no. 5,
21. pp. 2236–2257, 2020, doi:10.3837/TIIS.2020.05.021

22. M. Alam, S. Bhattacharya, S. Dutta, S. Sinha, D. Mukhopadhyay, and A. Chattopadhyay, __RATAFIA: Ransomware analysis using time and frequency informed autoencoders,‘‘ in Proc. IEEE Int. Symp.Hardw. Oriented Secur. Trust (HOST), May 2019, pp. 218–227, doi:10.1109/HST.2019.8740837

23. H. Zuhair and A. Selamat, __RANDS: A machine learning-based anti-ransomware tool for Windows platforms,‘‘ in Advancing Technology Industrialization Through Intelligent Software Methodologies, Tools and Techniques, vol. 318, 2019.

24. Y. A. Ahmed, B. Kocer, and B. A. S. Al-rimy,
25. __Automated analysis approach for the detection of high survivable ransomware,‘‘ KSII Trans. Internet Inf. Syst., vol. 14, no. 5, pp. 2236–2257, 2020, doi:10.3837/TIIS.2020.05.021

26. Sajad Homayon,Ali Dehghanaha,Marzeih Ahmadzedh,Sattar Hashemi,Raouf Khayami Know Ab- normal, Find Evil: Frequent Pattern Mining for Ransomware Threat Hunting and Intelligence https://ieeexplore.ieee.org/document/8051108.

27. A comprehensive Survey ransomware attack: A growing havoc cyber threat Tandon et al https://link.springer.com/chapter/10.1007/978-981-13-1274- 831.
28. 22] Yachya Abukar Ahmed, Barus Kocer, Bander Ali Saleh Alrimg Automated Analysis Approach for detection of high survivable ransomware https://www.researchgate.net/publication/341776333 Automated Analysis Approach for the Detection of High Survivable Ransomware.
29. [23] Manner Alam, Sarani Bhattacharya, Swastika Dutta, Sayan Sinha . RATAFIA: Ransomware Analysis using Time and Frequency Informed Autoencoder
30. https://ieeexplore.ieee.org/document/8740837.

31. 24] Kharraz et al[20], Shaukat and Ribeiro[24] A Digital DNA Sequencing Engine For ransomware detection using ML. https://ieeexplore.ieee.org/document/9121260.

32. Hiba Zuhair and Al Selamat RANDS: A Machine Learning Based Anti Ransomware Tool for win- dows platform https://www.researchgate.net/publication/340267182 RANDS a Machine Learning- Based Anti-Ransomware Tool for Windows Platforms.

33. Aurangzeb, S.; Aleem, M.; Iqbal, M.A.; Islam, M.A. Ransomware: A survey and trends. J. Inf. Assur. Secur. 2017, 6, 48–58.
34. Subedi, K.P.; Budhathoki, D.R.; Dasgupta, D. Forensic analysis of ransomware families using static and dynamic analysis. In Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 24 May 2018; pp. 180–185.

35. Alhawi, O.M.; Baldwin, J.; Dehghantanha, A. Leveraging machine learning techniques for windows ransomware network traffic detection. In Cyber Threat Intelligence; Springer: Cham, Switzerland, 2018; pp. 93–106

36. Andronio, N.; Zanero, S.; Maggi, F. HelDroid: Dissecting and Detecting Mobile Ransomware. In Research in Attacks, Intrusions, and Defenses; Bos, H., Monrose, F., Blanc, G., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 382– 404.

37. Sgandurra, D.; Mu ˜noz-Gonz ´alez, L.; Mohsen, R.; Lupu,
38. E.C. Automated dynamic analysis of ransomware: Benefits, limitations and use for detection. arXiv 2016, arXiv:1609.03020.

39. Ferrante, A.; Malek, M.; Martinelli, F.; Mercaldo, F.; Milosevic, J. Extinguishing ransomware- a hybrid approach to android ransomware detection. In Proceedings of the International Sym- posium on Foundations and Practice of Security, Nancy, France, 23–25 October 2017; Springer:
40. Berlin/Heidelberg, Germany, 2017; pp. 242–258.

41. Kara, I.; Aydos, M. Static and dynamic analysis of third generation cerber ransomware. In Proceed- ings of the 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), Ankara, Turkey, 3–4 December 2018; pp. 12–
42. 17.

43. Fernandez Maimo, L.; Huertas Celdran, A.; Perales Gomez, A.L.; Garcia Clemente, F.J.; Weimer, J.; Lee, I. Intelligent and dynamic ransomware spread detection and mitigation in integrated clinical environments. Sensors 2019, 19, 1114.

44. Alhawi, O.M.; Baldwin, J.; Dehghantanha, A. Leveraging machine learning techniques for windows ransomware network traffic detection. In Cyber Threat Intelligence; Springer: Cham, Switzerland, 2018; pp. 93–106.

45. L. Simonovich. (Jan. 15, 2020). Are Utilities Doing Enough to Protect Themselves From Cyberat- tack?. World Economic Forum. Accessed: Apr. 4, 2021.

46. Ramadhan A. M. Alsaid.Ransomware Detection using Machine and Deep Learning Approaches available on:https://www.researchgate.net/publication/365971703Ranso mwareDetectionusingMachineandDeepLearn

47. Chen, Z.G.; Kang, H.S.; Yin, S.N.; Kim, S.R. Automatic ransomware detection and analysis based on dynamic API calls flow graph. In Proceedings of the International Conference on Research in Adaptive and Convergent Systems, Krakow, Poland, 20–23 September 2017; pp. 196–201.
48. Hwang, J.; Kim, J.; Lee, S.; Kim, K. Two-Stage Ransomware Detection Using Dynamic Analysis and Machine Learning Techniques.Wirel. Pers. Commun. 2020, 112, 2597–2609.

49. Kharaz, A.; Arshad, S.; Mulliner, C.; Robertson, W.; Kirda, E. UNVEIL: A large-scale, auto- mated approach to detecting ransomware. In Proceedings of the 25th USENIX Security Symposium (USENIX Security 16), Austin, TX, USA, 10–12 August 2016; pp. 757–772.

50. Cabaj, K.; Mazurczyk, W. Using software-defined networking for ransomware mitigation: the case of cryptowall. IEEE Netw. 2016, 30, 14–20.
51. QuoIntelligence. (Jan. 18, 2022). Ransomware is Here to Stay and Other Cy- bersecurity Predictions for 2022. Accessed: Jan. 31, 2021. [Online]. Available:
52. https://quointelligence.eu/2022/01/ransomware-and-other- cybersecurity-predictions-for-2022

53. D. Golden and K. Norton. (2021). Defending Against Ransomware in
54. an Age of Emerging Technology. Deloitte. Accessed: Jan. 31, 2021.
55. [Online].Available:https://www2.deloitte.com/us/en/pages/ris k/articles/defending-against-
56. ransomware.html

57. APWG. (May 11, 2020). Phishing Activity Trends
58. Report in Q1 of 2020. Accessed: Apr. 4, 2021.
59. Q. Chen and R. A. Bridges, ―Automated behavioral analysis of mal- ware: A case study of Wan- naCry ransomware,‖ in Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2017, pp. 454– 460, doi: 10.1109/ICMLA.2017.0-119.

60. (May 22, 2017). WannaCry Ransomware Campaign Exploiting SMB Vulnerability. Accessed: Apr. 4, 2021. [Online]. Available: https://cert.europa.eu/static/SecurityAdvisories/2017/CERT- EU-SA2017-012.pdf [47] M. Akbanov, V. G. Vassilakis, and
61. M. D. Logothetis, ―WannaCry ran- somware: Analysis of infection, persistence, recovery prevention and prop- agation mechanisms,‖ J. Telecommun. Inf. Technol., vol. 1, no. 2019,pp. 113–124, Apr. 2019.

62. L. J. Trautman and P. Ormerod, ―Wannacry, ransomware, and the emerging threat to corpora- tions,‖ SSRN Electron. J., vol. 86, p. 503, Jan. 2018, doi:10.2139/ssrn.3238293.

63. S. Jones and T. Bradshaw. (May 14, 2017). Global Alert to Prepare for Fresh Cyber Attacks. Accessed: Apr. 4, 2021. [Online]. Available:https://www.ft.com/content/bb4dda38-389f-11e7-821a-
64. 6027b8a20f23

65. M. V. Liy. (May 15, 2017). Putin Culpa a Los Servicios Secretos de EE UU Por el Virus ‘Wan- naCry‘ Que Desencaden? el Ciberataque Mundial. Accessed: Apr. 4, 2021. [Online]. Available: https://elpais.com/internacional/2017/05/15/actualidad/149485 5826022843.html
66. S. K. Sahi, ―A study of wannacry ransomware attack,‖ Int. J. Eng. Res.Comput. Sci. Eng., vol. 4, no. 9, pp. 5–7, 2017.

67. R. Collier, ―NHS ransomware attack spreads worldwide,‖ Can. Med. Assoc. J., vol. 189, no. 22, pp. E786– E787, 2017.

68. JavaScript—MDN. (Feb. 18, 2022). JavaScript Language Resour
69. ces—JavaScript: MDN. Accessed: Apr. 4, 2021. [Online]. Available: https://developer.mozilla.org/enUS/docs/Web/JavaScript/Lang uage Resources

70. J. Gerend. (Mar. 3, 2021). Wscript. Microsoft Docs.
71. Accessed: Apr. 4, 2021. [Online]. Available: https://docs.microsoft.com/en-us/windows server/administration/windows-commands/wscript

72. T. McIntosh, A. S. M. Kayes, Y.-P.-P. Chen, A. Ng, and
73. P. Watters, ―Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions,‖ ACM Comput. Surv., vol. 54,no. 9, pp. 1–36, Dec. 2022, doi: 10.1145/3479393.

74. H. Oz, A. Aris, A. Levi, and A. S. Uluagac, ―A survey on ransomware: Evolution, taxonomy, and defense solutions,‖ ACM Comput. Surv., vol. 54, no. 11s, pp. 1–37, Jan. 2022, doi: 10.1145/3514229.

75. CIS Security. (2019). Fall 2019 Threat of the Quarter: Ryuk Ransomware. Accessed: Apr. 5, 2021.
76. [Online]. Available: https://www.cisecurity.org/white- papers/fall-2019-threat-ofthe-quarter-ryuk-
77. ransomware/

78. H. Ke, H. Wu, and D. Yang, ―Towards evolving security requirements of industrial inter- net: A layered security architecture solution based on data transfer techniques,‖ in Proc. Int. Conf. Cyberspace Innov. Adv. Technol., New York, NY, USA, Dec. 2020, pp. 504–511, doi:10.1145/3444370.3444620
79. Trend Micro. What is Ryuk Ransomware. Accessed: Apr. 5, 2021. [Online].Available: https://www.trendmicro.com/enus/what-is/ransomware/ryukransomware.html.
80. WannaCry Ransomware. (May 15, 2017). WannaCry Ransom ware—LogRhythm. Accessed: Apr.
81. 14, 2021. [Online]. Available: https://logrhythm.com/blog/wannacry-ransomware/
82. A. Kujawa. (Jan. 8, 2019). Ryuk Ransomware Attacks Businesses Over
83. the Holidays. Malwarebytes Labs. Accessed: Apr. 14, 2021. [Online].
84. Available:https://blog.malwarebytes.com/cybercrime/malwar e/2019/01/ryuk-ransomware-
85. attacks-businesses-over-the-holidays/
86. R. Nimbalkar. (Jul. 13, 2021). Decision Tree
87. Algorithms-Machine Learning. Accessed: Apr. 14, 2021. [Online]. Available: https://medium.com/appengine-ai/decision-tree-algorithms-machine-learning- 9e2e8cadfcae[63] S. India. (Jul. 4, 2020). Hands-on Training With Machine Learn-ing Algorithms: Decision Tree and Random Forest. Springboard Blog.Accessed: Apr. 14, 2021.