WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Ransom Prediction Using ML Algorithm

Kanakam Maruthi Narasimha Rao[1], Konda Saiteja[2], Kamini Vasu[3], Dr. Kumar P[4], Dr. K. S. Ramanujam[5], Dr. V. B. Ganapthy[6]

[1,2,3,5,6] Department of CSE, Dr. MGR Educational Institute and Research Institute, Chennai, Tamilnadu, India
[4]Department of MECH, Dr. MGR Educational and Research Institute Chennai,Tamilnadu, India
[1]maruthiyadav987@gmail.com, [2]saikonda55134@gmail.com, [3]vasukamini9@gmail.com, [4]rajkumar.mech@drmgrdu.ac.in, [5]ramanujam.cse@drmgrdu.ac.in, [6]ganapathy.cse@drmgrdu.ac.in
DOI : https://doi.org/10.55248/gengpi.6.0425.1517

ABSTRACT-

As wi-fi conversation is growing, our region is dealing with many security threats. Predictive framework for ransomware malware detection and system threat detection. Many ML techniques have been used within the beyond to improve the accuracy of malware detection and prevention systems. In this examine, a technique become proposed to perceive the effectiveness of a ransomware system the usage of principal aspect evaluation (PCA) and random type techniques. PCA can reduce and arrange the know-how length, even as random wooded area can assist in classifier. The results display that the proposed approach plays better in phrases of accuracy than different techniques which include SVM, NB, and DT. The results of the proposed method show that the accuracy price (%) is 96.78%, the error price (%) is 0.21%, and the execution time (minimal) is 3.24 minutes..

Keywords: Online Security Risks, Ransom Ware, ML, RF, PCA, Outperforms

INTRODUCTION

Network protection threats have elevated dramatically due to the increase of Wi-Fi communication. Predictive marketing and advertising tools are very critical for detecting laptop threats and viruses. Previous research have used diverse machine studying (ML) strategies to detect ransomware, improving detection electricity and accuracy. This examine proposes a powerful sales forecasting strategy based on the random woodland method and main factor evaluation (PCA). The random woodland method improves the general magnificence overall performance, while PCA allows in shaping the tables via restricting the scale of the records. The results display that the proposed approach outperforms existing techniques together with aid vector system (SVM), neural base and tree selection in terms of accuracy and efficiency. The machine is expected to run for 3.24 minutes with 96.78% accuracy and 0.21% mistakes.

Malware is defined as any action that jeopardizes the confidentiality, availability, or integrity of data or computer sources. Cybercriminals circumvent authentication or authorization procedures by taking advantage of errors or defects in computer design. Because of the rapid expansion of community services and the increasing demand for safe records transit, network security is more crucial than ever. An important technique for identifying assaults is ransomware prediction frameworks that monitor a variety of community sports. These structures can accurately identify dangers, conduct rapid research, and lower fake positives. By identifying volatile anomalies, displaying information about suspicious activity, warning administrators, and stopping harmful activity, ransomware prediction systems help safeguard networks. There are two main forms of predictive ransomware structures: host-based and network-based. The goal of community-based ransomware detection is to identify odd and invalid hobbies by analyzing network visitors. However, a Host Intrusion Detection System (HIDS) tracks intrusions and examines device logs, report integrity, and procedure behavior to identify harmful hobbies on a group. Because they provide protection against computer hacking, data theft, and illegal access, both types are crucial for maintaining cybersecurity. Creating a reliable and efficient ransomware attack prediction system is essential to protecting virtual property and maintaining network integrity in light of the increasing complexity of cyberthreats.

RELATED WORK

We recommend a signature-based totally intrusion detection device design, which incorporates strategies to locate all of the above threats. According to our experimental results, IDS cannot detect any of the nine assaults, whilst AP is sensitive to 8 of them. The primary disadvantage of this approach is that none of them may be detected with the aid of IDS.

Then, using the compressed characteristic area, we employ the following device learning techniques: LR, kNN,SVM, ANN and DT. Our analysis took into account both binary and multi-magnificence classes. The findings demonstrate that the experimental accuracy of the DT-like binary schema category

technique is increased from 88.Thirteen to 90.85% with the aid of the XGBoost-based row selection approach. One disadvantage is that it typically requires careful calibration, particularly when the size exceeds the number of samples.[2].

We in comparison five device mastering algorithms: logistic regression, tree pruning, random wooded area, XGB, and artificial neural networks. We evolved a brand new artificial neural network structure that outperformed the other algorithms, accomplishing ~ninety nine.2% accuracy. We executed our experiments on a Linux system going for walks Ubuntu 21.04 with 32GB RAM and 8GB Nvidia GTX.

We used a 1080 GPU to run our decentralized schooling set of rules for quicker outcomes. Only approximately 81% accuracy became accomplished on the ADFA intrusion detection dataset. There are some drawbacks to anomaly detection [3].

Use Attack Intent Analysis (AIA) to be expecting assaults. Let us begin through discussing the structure of this machine. Research in the field of security device improvement specializes in tracking and analysing attacks without considering the real methods to protect against those assaults. This process is tough to govern [4]. The proposed KDE-HMM method works well. For better results, the proposed KDE-HMM technique/system combines the advantages of statistics and probabilistic methods. Experimental validation has showed the effectiveness of the proposed KDE-HMM method in detecting recorded threats with 98% accuracy. The fundamental drawback in their approach is that it makes use of a performance threshold to pick out the first-class features. Therefore, to hit upon intrusions, cutting-edge intrusion detection structures (IDS) based totally on random models consisting of HMM only examine the conduct of the attacker and the preliminary infection vectors consisting of device calls, machine movements, and signatures. In addition, these parameters aren't sufficient to stumble on intrusions given the attacker's want to advantage and control the gadget [5].

EXISTING SYSTEM

Ahmed Iftikhar et al. Along with Random Jump, SVM, and Convolutional Neural Network (CNN), they also compared a number of other techniques. Convolutional Neural Networks perform better than any other approaches, according to the authors' findings. Here, B. Riaz et al. optimized the dataset using a rule-based, fully feature-choice method. They verified a dynamic rise in the so-called IDS using the KDD dataset.

Disadvantages

When taking walks on the Internet, a lot of malicious sports have an impact on the structures. Facts leaking as a result of computer breach is the main problem here. The accuracy, detection cost, and false alarm rate may all be improved based on the available data. The previously employed procedures can be replaced by other methods like SVM and Naive Bayes. The study also makes the case that there are methods for enhancing the dataset. Enhance the input's satisfaction for the machine's benefit.

PROPOSED SYSTEM

The ransomware prediction machine's primary reason is optimizing the compromised system. This system is capable of detecting malware. The proposed system is an attempt to address the last problems from previous studies. Two methods are used to create the proposed system: Uncertainty Forest and Principal Component Analysis. PCA enhances the quality of the dataset and provides it with the required first-class quality by minimizing the size of the information set. The RF set of rules, which performs better than SVM, can then be used to locate the Trojan horse.

Advantages:

The proposed method has a minimum errors price of zero.21%. Moreover, it has performed a lot better accuracy compared to the preceding method. Compared to other techniques, the effect takes less time.

SYSTEM ARCHITECTURE

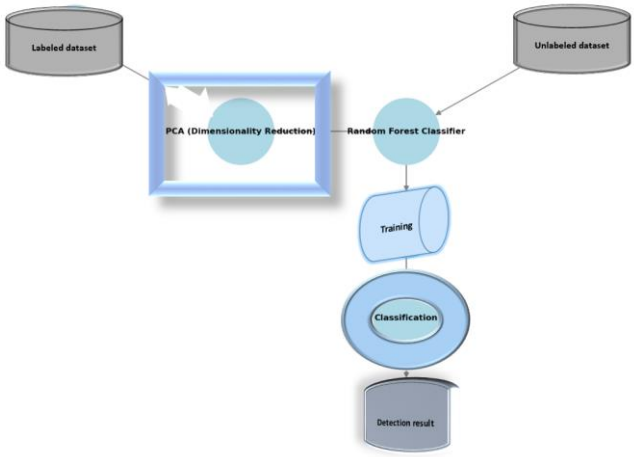

Fig 1: System Architecture

SYSTEM METHODOLOGY

RF

RF, a well-known gadget study technique, is a component of the observe method. It can be applied to regression problems and every class in academic packages. It is entirely built on the idea of ensemble mastery, which is a way to use several classifiers to solve a complex problem and enhance version performance. As the name suggests, a RF is a classifier that includes many selection trees for certain dataset units and improves forecast accuracy by expanding the dataset. A random woodland forecasts the outcome by taking the predictions from each decision tree and figuring out how many people vote for them, as opposed to depending on just one decision tree. The more shrubs there are in the wooded area, the more dispersed it is and the more effective it is in preventing them. The well-known system mastering method RF was developed by Leo Breiman and Adele Cutler. It combines the outputs of numerous decision trees to provide an unmarried outcome. Its adaptability and simplicity of usage have encouraged its use since it might resolve type and regression issues. This post describes the RF algorithm's operation, how it varies from other algorithms, and how to use it.

Algorithm

Step:1 The first step is to gather the data you want to use to train the RF.

Step: 2 The problem that RF is meant to solve must then be specified. In this case, binary classification is the problem.

Step: 3 To evaluate how well the RF model performs, the data must be separated into training and test sets.

Step: 4 The RF model's performance on the test set can be evaluated after it has been trained.

Step: 5 If the RF model's performance is poor, you may need to change the hyper parameters to improve the results.

Step: 6 You can use the RF model to forecast new, untested data after you are satisfied with its performance.

SYSTEM MODULES

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on test set
- Saving the Trained Model

Module Descriptions

Data Collection:

In the construction of the ML version of the statistics series, this is the first actual step. The better the version, the more important this step is the more accurate the information we will get, and the more likely it is to be known, which depends on how well our version performs. There are numerous fact series tactics, such as manual intervention and textual material scraping.

Datasets:

The dataset contains 125974 unique data points. There are 42 columns in the dataset.

Num_file_creations	Number of file creation operations	continuous
Num_shells	Number of shell prompts	continuous
Num_access_files	Number of operations on access control files	continuous
Num_outbound_cmds	Number of outbound commands in an ftp session	continuous
Is_hot_login	1 if the login belongs to the"hot"list;0 otherwise	discrete
Serror_rate	% of connections that have "SYN"errors	continuous
Same_srv_rate	% of connections to the same service	continuous
Diff_srv_rate	% of connections to different services	continuous
Srv_count	Number of connections to the same services as the current connection in the past two seconds	continuous
Srv_serror_rate	% of connections that have "SYN"errors	continuous
Srv_rerror_rate	% of connections that have "REJ"errors	continuous
Srv_diff_host_rate	% of connections to different hosts	continuous
Num_root	Number of "root"accesses	continuous

Feature name	Description	Type
Duration	Length(number of seconds)of the connection	continous
services	Network services Service on the destination,e.g., http,etc.	discrete
Src_bytes	Number of data bytesfrom source To destination.	continous
Dst_bytes	Number of data bytes from destination to source	continous
Flag	Normal or error status of the connection	discrete
Land	1 if connection is from/to the same host/port;0 otherwise	discrete
Wrong_fragment	Number of "wrong"fragments	continous
urgent	Number of urgent packets	comtiuous
Hot	Number of "hot"indicators	continuous
Num_failed_logins	Number of failed login attempts	continuous
Logged_in	1 if successfully logged in;0 otherwise	discrete
Num_compromised	Number of "compromised" conditions	continuous
Root_shell	1 if root shell is obtained;0 otherwise	discrete
Su_attempted	1 if "su root"command attempt;0 otherwise	discrete

Data Preparation:

The statistics are being converted by us. Eliminate several columns and statistics that are absent. We will start by compiling a list of column names that we must maintain. After that, we can remove or delete every column except the ones we must keep. Lastly, we remove or discard rows from the dataset that have missing values. Separate the series and look at each set.

Model Selection:

One method for reducing the scale of an information collection is principal thing evaluation. Main aspect analysis provides the intended outcomes and is one of the only and accurate ways to reduce the dimensions of information. This technique reduces a dataset's capabilities into a fixed number of capabilities known as primary components. All of the entered statistics are regarded as a dataset with numerous properties and dimensions in this procedure. The facts elements are compressed by this method, which arranges the statistics factors along an unmarried axis. By converting the record points into axes, the principal components are computed. With PCA, the following processes can be finished:

1. Take all of the dimensions of the dataset.

2. Determine the median vector of each measurement d.

3. Determine the general covariance matrix of the dataset.

4. Determine the eigenvalues (v1, v2,,.. V3,.. Vd) and eigenvectors (e1, e2, e three .....Ed).

5. To attain the d*n=M matrix, the eigenvalues are sorted in ascending order and the n eigenvectors with the highest eigenvalue are decided on.

6. Create a brand new subject version the use of this M shape.

7. The principal components are the sample space.

The first step includes building a wooded area the usage of the provided dataset, and the second one step involves creating a categorical prediction.

Saving the Trained Model:

The first step is to store it in a.H5 or.Pkl library report ALEX once you are positive that you have tested and skilled the version enough to move it to production. Verify that Pickle is configured in your surroundings. After that, the method may be imported and the version added to a.Pkl record.

RESULTS AND DISCUSSION

Fig 2: Home Page

Figure 2 displays the application's home page, where users can click Login to begin using it.

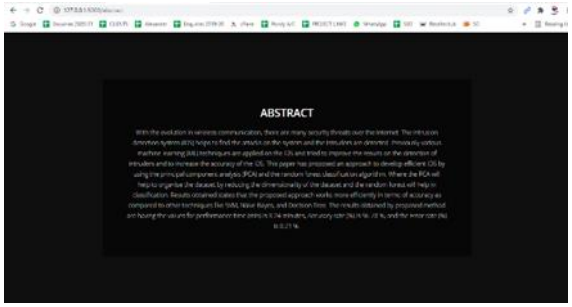

Fig 3: Abstract Page

Fig 3 shows the Abstract Page where the User can know about the application,

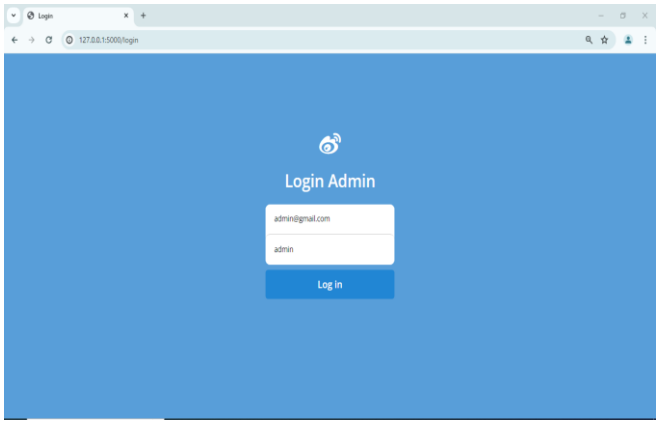

Fig 4: Login Page

The login page is displayed in Figure 4, where users can enter their ID and password to access the application.

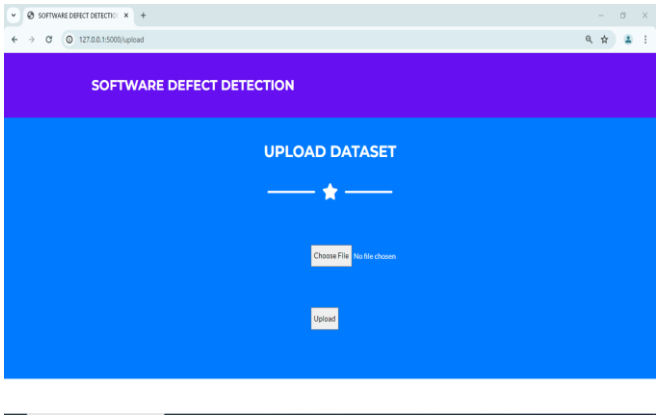

Fig 5: Upload Page

The Upload Page is displayed in Fig. 5, where the user can select and upload the dataset file.

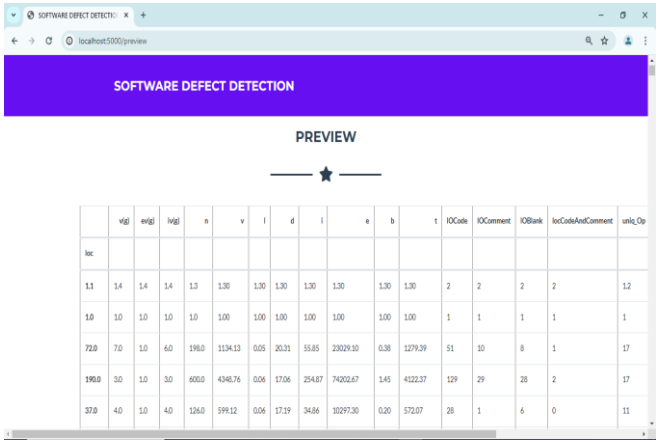

Fig 6: Preview Page

The Preview Page, where the user can examine a preview of the dataset they uploaded, is depicted in Figure 6.

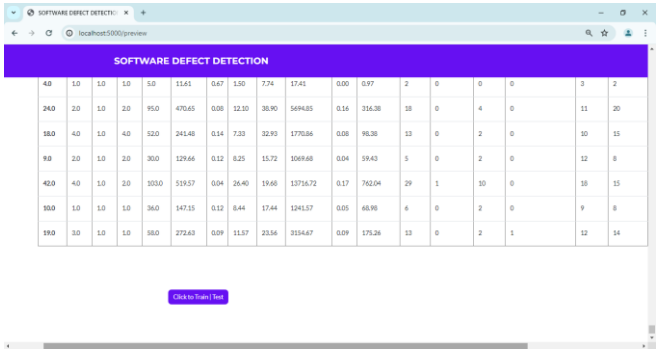

Fig 7: Train/Test Page

The Train/Test Page, where the user clicks to train or test the uploaded dataset, is displayed in Fig 7.

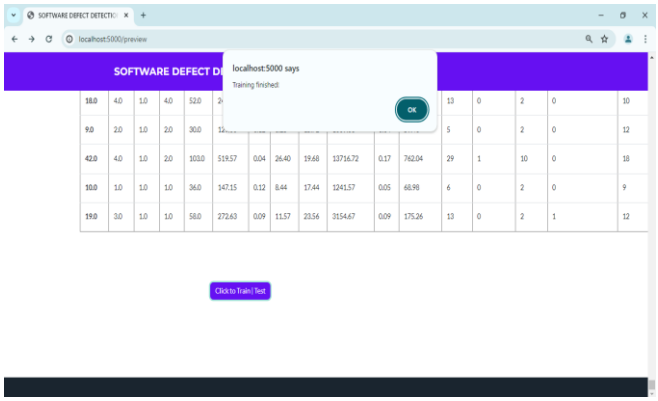

Fig 8: Training Complete Page

The Training Complete Page, where the user is informed that the training is finished, is displayed in Fig. 8.

Fig 9: Predict Label Page

The Predict Label Page, seen in Fig. 9, allows the user to enter information and determine whether the program is ransomware by selecting the Predict Label button.

CONCLUSION

Protection concerns are expanding along with the unanticipated growth in the number of businesses on the Internet. The difficulty of identifying Internet intruders is effectively resolved by the suggested method. The suggested algorithm outperforms previously employed algorithms such as SVM, DT and NB. The detection charge and false error fee can be significantly increased with the suggested method. The Science Discovery dataset is the one being used here. The results of using the suggested method are as follows: level accuracy (%) = 96.78%, level blunders (%) = 0.21%, and execution time (minimum) = 3.24 minutes.

REFERENCES

[1] JafarAbo Nada; Mohammad Rasmi Al-Mosa, 2018 International Arab Conference on Information Technology (ACIT), A Proposed Wireless Intrusion Detection Prevention and Attack System

[2] Kinam Park; Youngrok Song; Yun-Gyung Cheong, 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigData Service), Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm

[3] S. Bernard, L. Heutte and S. Adam "On the Selection of Decision Trees in RFs" Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009, 978-1-4244-3553-1/09 /2009 IEEE

[4] Tesfahun, D. Lalitha Bhaskari, "Intrusion Detection using RFs Classifier with SMOTE and Feature Reduction" 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies, 978-0-4799-2235-2/13 2013 IEEE

[5] Le, T.-T.-H., Kang, H., & Kim, H. (2019). The Impact of PCA-Scale Improving GRU Performance for Intrusion Detection. 2019 International Conference on Platform Technology and Service (PlatCon). Doi:10.1109/platcon.2019.8668960

[6] Anish Halimaa A, Dr.K.Sundarakantham: Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) 978-1-5386-9439-8/19/2019 IEEE "MACHINE LEARNING BASED INTRUSION DETECTION SYSTEM."

[7] Mengmeng Ge, Xiping Fu, Naeem Syed, Zubair Baig, Gideon Teo, Antonio Robles-Kelly (2019). Deep Learning-Based Intrusion Detect ion for IoT Networks, 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC), pp. 256-265, Japan.

[8] R. Patgiri, U. Varshney, T. Akutota, and R. Kunde, "An Investigation on Intrusion Detection System Using Machine Learning"978-1-5386-9276-9/18/ c2018IEEE.

[9] Rohit Kumar Singh Gautam, Er. Amit Doegar; 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) "An Ensemble Approach for Intrusion Detect ion System Using Machine Learning Algorithms."

[10] Kazi Abu Taher, Billal Mohammed Yasin Jisan, Md. Mahbubur Rahma, 2019 International Conference on Robot ics, Electrical and Signal Processing Techniques (ICREST)"Network Intrusion Detect ion using Supervised Machine Learning Technique with Feature Selection."

[11] L. Haripriya, M.A. Jabbar, 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)" Role of Machine Learning in Intrusion Detection System: Review"

[12] Nimmy Krishnan, A. Salim, 2018 International CET Conference on Control, Communication, and Computing (IC4) " Machine Learning-Based Intrusion Detect ion for Virtualized Infrastructures"

[13] Mohammed Ishaque, Ladislav Hudec, 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) "Feature extract ion using Deep Learning for Intrusion Detection System."

[14] Aditya Phadke, Mohit Kulkarni, Pranav Bhawalkar, Rashmi Bhattad, 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)"A Review of Machine Learning Methodologies for Network Intrusion Detection."

[15] Iftikhar Ahmad , Mohammad Basheri, Muhammad Javed Iqbal, Aneel Rahim, IEEE Access (Volume: 6) Page(s): 33789 – 33795 "Performance Comparison of Support Vector Machine, RF, and Extreme Learning Machine for Intrusion Detection."

[16] B. Riyaz, S. Ganapathy, 2018 International Conference on Recent Trends in Advanced Computing (ICRTAC)" An Intelligent Fuzzy Rule-based Feature Select ion for Effective Intrusion Detection."