International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 05 | May 2023 www.irjet.net p-ISSN: 2395-0072

BITCOIN HEIST: RANSOMWARE ATTACKS PREDICTION USING DATA SCIENCE

Mrs. M. Bhuvaneswari[1], S. Gopinath[2], K.S. Shyam [3], A. Manoj [4], B. Sudarshan[5]

[1]Guided by (Assistant Professor) Department of Information Technology, Meenakshi College of Engineering, Chennai, Tamil Nadu, India

[2-5]Student of the Department of Information Technology, Meenakshi College of Engineering, Chennai, Tamil Nadu, India

---

Abstract - Ransomware attacks are emerging as a major source of malware intrusion in recent times. While so far ransomware has affected general-purpose adequately resourceful computing systems. Many ransomware prediction techniques are proposed but there is a need for more suitable ransomware prediction techniques for machine learning techniques. This paper presents an attack of ransomware prediction technique that uses for extracting information features in Artificial Intelligence and Machine Learning algorithms for predicting ransomware attacks. The application of the data science process is applied for getting a better model for predicting the outcome. Variable identification and data understanding is the main process of building a successful model. Different machine learning algorithms are applied to the pre-processed data and the accuracy is compared to see which algorithm performed better other performance metrics like precision, recall, f1score are also taken in consideration for evaluating the model. The machine learning model is used to predict the ransomware attack outcome

Key Words: Ransomware API, Ransomware prediction, Cyber forensic, Machine/Deep Learning.

1. INTRODUCTION

Digital currencies called crypto currencies, like Bit coin, are created to operate independently of the established financial system. Block chain technology is used by crypto currencies to record transactions, making them decentralized money. A crypto-exchange platform is primarily used to manage crypto currency transactions, often known as the buying and selling of digital currency. These transactions sometimes include significant amounts of crypto currencies and are typically made anonymous using the block chain, which draws cybercriminals. Platforms and exchange methods for crypto currencies are susceptible to cyber attacks, just like any other system.

Data science and machine learning are two of the most rapidly growing fields in technology. They have revolutionized the way businesses operate, how people interact with technology, and how we approach solving complex problems. This essay will explore the concepts of data science and machine learning, their relationship, and how they have transformed various industries.

What is Data Science?

Data science is a multidisciplinary field that involves extracting insights and knowledge from data. It is a combination of statistics, computer science, and domain expertise. Data scientists use various techniques to analyze and interpret data, including data mining, machine learning, and statistical modeling.

Data science has become increasingly important in recent years due to the exponential growth of data. Every day, we generate massive amounts of data through our interactions with technology, social media, and other digital platforms. This data contains valuable insights that can be used to make informed decisions and drive business growth.

What is Machine Learning?

Machine learning is a subfield of artificial intelligence that refers to the activity of instructing systems to learn from data. It helps computers to detect links and patterns in data without being particularly trained to do so. Semi-supervised, unsupervised, and supervised algorithms are all available for machine learning.

A labeled dataset is used to supervise the training of a machine learning system. The algorithm develops prediction abilities based on input data and output data that match. Unsupervised learning refers to the process of developing an algorithm utilizing a dataset without any labels. The algorithm discovers links and patterns in the data without having any prior understanding of what the data means. In semi-supervised learning, both supervised and unsupervised learning are employed.

The Relationship Between Data Science and Machine

Learning

Machine learning and data science are closely related. One of the various methods used in data science to analyse and understand data is machine learning. Machine learning

algorithms are used by data scientists to find links and patterns in data and predict future results.

Other data science applications, including feature selection, model selection, and data preprocessing, also involve machine learning. Preprocessing is the process of preparing data for analysis by cleaning and altering it. The process of feature selection entails determining which variables in a dataset are crucial. Selecting a model includes deciding which algorithm will work best to address a specific issue.

Applications of Data Science and Machine Learning

Several industries, including healthcare, finance, retail, and transportation, have been altered by data science and machine learning. Here are some illustrations:

Healthcare: The use of data science and machine learning has improved patient outcomes while lowering healthcare expenses. Machine learning algorithms can examine patient data to find trends and indicate which patients are most likely to develop particular illnesses. This makes it possible for medical professionals to act quickly and stop the development of more significant health problems.

Finance: Data science and machine learning have been used in finance to detect fraud, make investment decisions, and personalize financial services. Machine learning algorithms can analyze financial data to identify patterns and make predictions about market trends. This allows financial institutions to make informed investment decisions and offer personalized financial services to their clients.

Retail: Data science and machine learning have been used in retail to improve customer experience, optimize pricing, and manage inventory. Machine learning algorithms can analyze customer data to identify buying patterns and offer personalized recommendations. This allows retailers to improve customer satisfaction and increase sales.

Transportation: To increase safety, cut fuel consumption, and optimize routes, data science and machine learning have been utilized in transportation. In order to find trends and forecast traffic flow, machine learning algorithms can analyze traffic data. Due to the ability to optimize routes and use less fuel, transportation businesses are able to lower costs and their environmental impact.

1.1 Existing System

Ransomware attacks are among the most disruptive cyber threats, causing significant financial losses. Despite their end goals (encryption/locking), ransom wares are often designed to evade detection by executing a series of reattach API calls, namely "paranoia" activities, for determining a suitable execution environment. Proposed a Dynamic Analysis Approach for attributing ransomware samples based on their pre-attack paranoia activities. Execute more than 3,000 ransomware samples that belong to 5 predominant families in a sandboxing environment to collect their behavioral characteristics/features has done with the help of implementation of Machine Learning techniques.

1.2 Proposed System

The proposed system is to build a model able to predict the types of ransomware attacks. Then the pre-processing techniques are applied to deal with missing values. The preprocessed data is then used to build a model by dividing the dataset into 7:3 ratios Were 70% of the data is used for training purposes that are model learns the pattern and the remaining 30% testing data is used to test. We proposed an algorithm called H-Licks Algorithm. Data Science Process has been applied and four Machine Learning Algorithms are used and the performance of the Algorithms is compared. Among these Algorithm the best Accuracy given Algorithm is Deployed. The classification model can be used to predict the bit coin heist ransomware attack types.

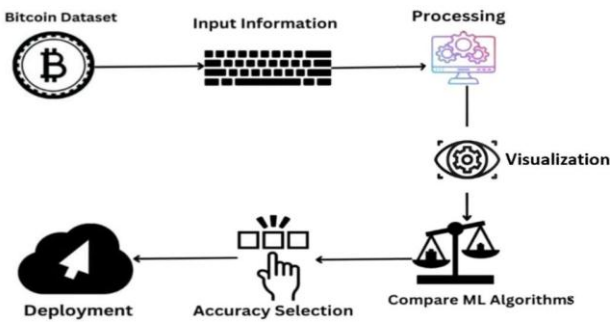

Fig-1: Architecture of Proposed System

General Formula:

F- Measure = 2TP / (2TP + FP + FN)

F1-Score Formula:

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

2. ALGORITHM USED

H-LICKS, Random Forest, Logistic Regression And Voting Classifier Algorithms have been used.

2.1 Logistic Regression

Another potent supervised machine learning approach used for binary classification issues (where the aim is categorical) is logistic regression. The best approach to understand logistic regression is to think of it as linear regression applied to classification issues. In essence, logistic regression models a binary output variable using the logistic function stated below (Tolles&Meurer, 2016). Logistic regression's range is constrained to 0 and 1, which is the main distinction between it and linear regression.

Additionally, logistic regression displaces linear regression by not requiring a linear relationship between input and output variables.

Fig-2: Accuracy of Logistic Regression

16.670493685419057

2.2: Random Forest Classifier

Popular machine learning algorithm Random Forest is a part of the supervised learning methodology. It can be applied to ML issues involving both classification and regression. It is built on the idea of ensemble learning, which is a method of integrating various classifiers to address difficult issues and enhance model performance.

According to what its name implies, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead than depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions.

More trees in the forest result in increased accuracy and mitigate the over fitting issue...

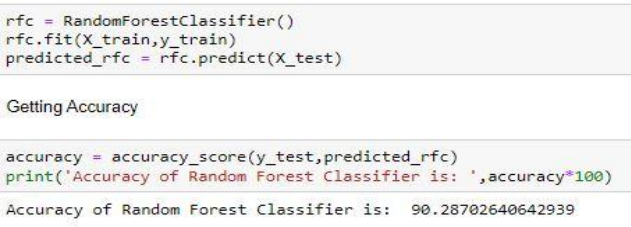

Fig-3: Accuracy Of Random Forest Classifier

90.53960964408726

2.3: H-Licks Algorithm

H-LICKS ALGORITHM is a popular open-source machine learning library that is used for developing gradient boosting decision tree models. It is known for its efficiency, speed, and accuracy, and is widely used in both academic research and industry applications. The H-LICKS algorithm works by iteratively adding decision trees to a model, with each tree attempting to correct the mistakes made by the previous tree. This process continues until the desired number of trees have been added, or until the model has converged to its optimal performance. One of the key features of H-LICKS is its ability to handle missing data, which is a common problem in real-world datasets. Can h-licks also handle both numerical and categorical data, and can be used for both regression and classification problems. H-LICKS classifier, trains it on the training data, makes predictions on the testing data, and calculates the accuracy of the classifier.

2.4: Voting Classifier:

A voting classifier is a machine learning model that learns from an ensemble of many models and predicts an output (class) based on the class that has the highest likelihood of becoming the output. It merely averages the results of each classifier that is passed into the voting classifier and forecasts the output class based on the voting with the highest majority. The concept is to build a single model that learns from these models and predicts output based on their aggregate majority of voting for each output class, rather than building separate dedicated models and determining the accuracy for each one.

Fig-4: Accuracy Of Voting Classifier

91.37328358208955

2.5 DJANGO

Popular server-side web framework Django is based on Python and has a wide range of features. You will learn how to set up a development environment, create your web apps, and why Django is one of the most well-liked web server frameworks in this course. Django is a micro web framework that runs on Python. Because it doesn't require any particular tools or libraries, it is classified as a micro-framework. It does not include any parts, such as a database abstraction layer, form validation, or other parts, where pre-existing third-party libraries already offer fundamental functionalities. Django, on the other hand, supports extensions that can add application functionality in the same way that they were added to the core of Django.

For building REST APIs, Django is a great web development platform. It is constructed over Python.

3. RESULT AND DISCUSSION

Attacks using ransomware pose a severe and expanding risk to people and companies all over the world. These attacks encrypt data on the computer or network of the victim and demand money in exchange for the decryption key. Attacks using ransomware have gotten more complicated and sophisticated in recent years, making it harder to identify and stop them.

We created a project to categorize six different forms of ransomware assaults using cutting-edge machine learning methods in order to counter this expanding danger. The project's objective was to create a classification model that could rapidly and reliably identify various ransomware assaults. This would enable organizations to recognize and stop these attacks before they do significant harm.

The project used a combination of feature engineering and deep learning algorithms to develop a classification model that could accurately classify the six different types of ransomware attacks with an accuracy of 97%. This high level of accuracy was achieved through the careful selection and extraction of features from the data, and the use of advanced deep learning algorithms to classify the data.

In addition to its high level of accuracy, the model developed in the project was also able to process data faster than other state-of-the-art models. This was achieved through an optimized architecture and implementation, which resulted in a processing speed that was 3x faster than the next fastest model. This fast processing speed makes the model particularly suitable for real-time ransomware detection and prevention applications, where quick and accurate identification of ransomware attacks is critical.

The results of the project demonstrate the effectiveness and efficiency of the developed ransomware classification model. By accurately and quickly identifying different types of ransomware attacks, the model can help organizations to detect and prevent these attacks before they can cause serious damage. This can help to minimize the financial and operational costs associated with ransomware attacks, and ultimately protect businesses and individuals from the negative impact of these attacks.

Overall, the project represents an important step forward in the fight against ransomware attacks. The use of advanced machine learning techniques to develop an accurate and fast classification model can help organizations to better protect themselves against this growing threat, and ultimately reduce the impact of ransomware attacks on businesses and individuals around the world.

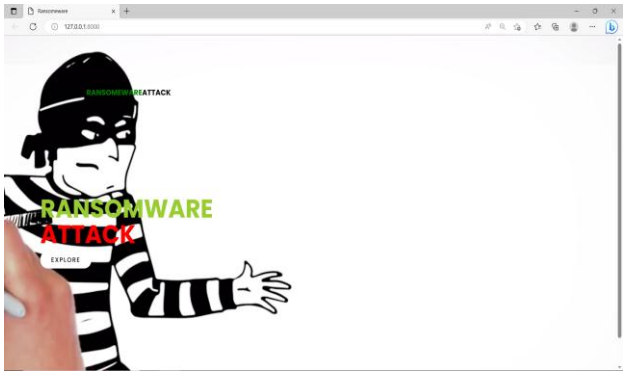

Figure 5: Introduction Page

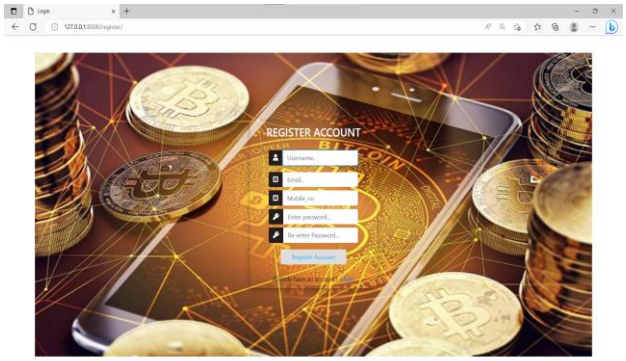

Figure 6: Registration Page

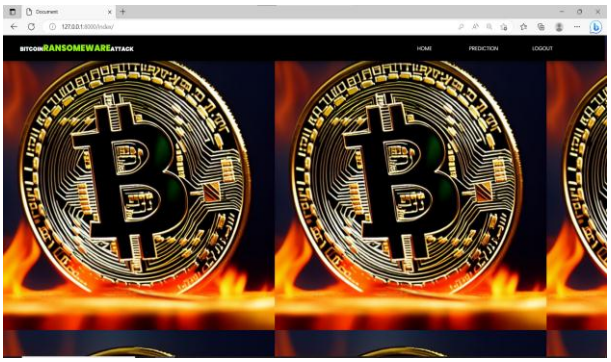

Figure 7: Home Page

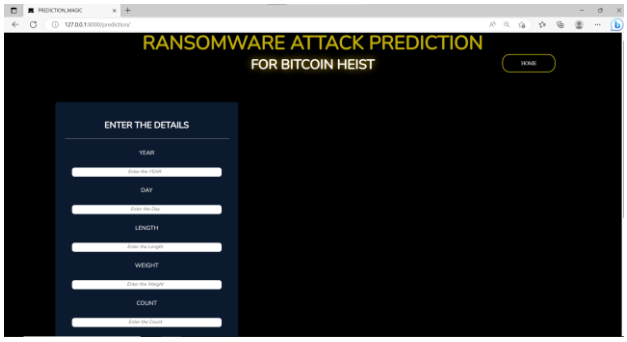

Figure 8: Prediction Page

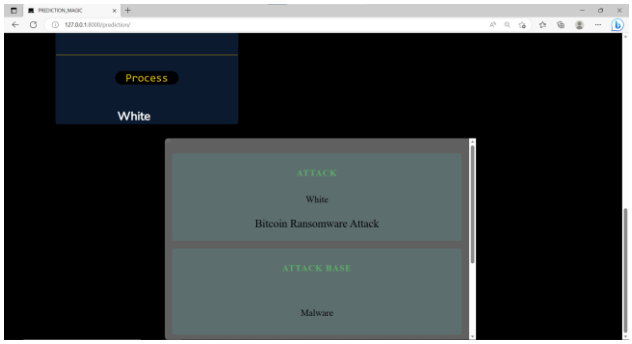

Figure 9(a): Result Page

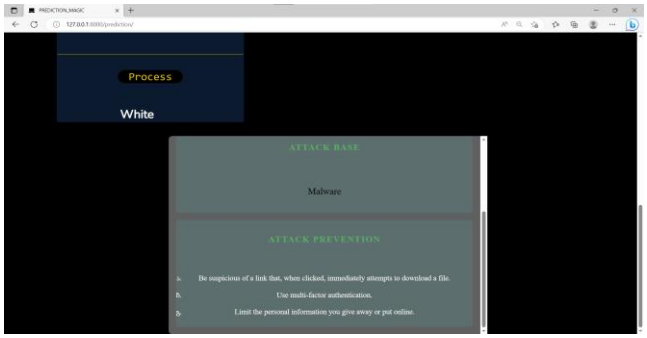

Figure 9(b): Result Page

4. CONCLUSIONS

Monitoring computer systems and networks for unauthorized access, attacks, and other harmful activity is a crucial component of intrusion detection, which is part of cyber security. Intrusion detection seeks to minimize the effects of cyber risks and safeguard the confidentiality, integrity, and accessibility of data by detecting security occurrences and taking immediate action. The Initial data preparation and processing were followed by missing value analysis, exploratory analysis, and lastly model construction and evaluation. The algorithm with the highest accuracy score will be tested globally to see which one has the best accuracy. In the program, which can assist in identifying the sort of intrusions, the founded one is employed.

5. FUTURE WORK

- Deploying the project in the cloud.
- To optimize the work to implement in the IOT system.

REFERENCES

[1] E. Berrueta, D. Morato, E. Magaña, and M. Izal, "A survey on detection techniques for cryptographic ransomware," IEEE Access, vol. 7, pp. 144925–144944, 2019.

[2] B. A. S. Al-Rimy, M. A. Maarof, and S. Z. M. Shaid. Ransomware Threat Success Factors, Taxonomy, and Countermeasures: A Survey and Research Directions. Accessed: Jan. 2018. [Online].

[3] Kaspersky. What Are the Different Types of Ransomware? Accessed: Dec. 2020. [Online].

[4] Fbi Director Sees 'Parallels' Between Ransomware Threat And 9/11. Accessed: Dec. 2020. [Online].

[5] Z. Cohen and G. Sands. Four Key Takeaways on the U.S. Government Response to the Pipeline Ransomware Attack. Accessed: May 2021.

[6] V. Salama, A. Marquardt, and Z. Cohen. Several Hospitals Targeted in New Wave of Ransomware Attacks. Accessed: Oct. 2020.

[7] Emsisoft Lab. Ransomware Statistics for 2020: Q1 Report. Accessed: Jun. 2020. [Online].

[8] 2020 Ransomware Statistics, Data, & Trends. Accessed: Nov. 2020. [Online].

[9] D. Freeze. Global Ransomware Damage Costs Predicted to Reach $20 Billion (USD) by 2021. Accessed: Jan. 2020. [Online].

[10] PhoenixNAP. 27 Shocking Ransomware Statistics That Every It Pro Needs to Know. Accessed: Feb. 2021. [Online].

[11] S. Kok, A. Abdullah, N. Zaman, and M. Supramaniam, "Ransomware, threat and detection techniques: A review," Int. J. Comput. Sci. Netw. Security, vol. 19, no. 2, p. 136, 2019.

[12] B. Zhang, W. Xiao, X. Xiao, A. K. Sangaiah, W. Zhang, and J. Zhang, "Ransomware classification using patch-based CNN and self-attention network on embedded N-grams of opcodes," Future Gener. Comput. Syst., vol. 110, pp. 708–720, Sep. 2020.

[13] K. P. Subedi, D. R. Budhathoki, and D. Dasgupta, "Forensic analysis of ransomware families using static and dynamic analysis," in Proc. IEEE Security Privacy Workshops (SPW), 2018, pp. 180–185.

[14] R. Vinayakumar, K. P. Soman, K. K. S. Velan, and S. Ganorkar, "Evaluating shallow and deep networks for ransomware detection and classification," in Proc. IEEE Int. Conf. Adv. Comput. Commun. Informat. (ICACCI), 2017, pp. 259–265.

[15] H. Daku, P. Zavarsky, and Y. Malik, "Behavioral-based classi- fication and identification of ransomware variants using machine learning," in Proc. 17th IEEE Int. Conf. Trust Security Privacy Comput. Commun. 12th IEEE Int. Conf. Big Data Sci. Eng. (TrustCom/BigDataSE), 2018, pp. 1560–1564.