Available online at sjuoz.uoz.edu.krd

Science Journal of University of Zakho

Vol. 10, No. 1, pp. 5– 10, March-2022

p-ISSN: 2663-628X
e-ISSN: 2663-6298

BITCOIN RANSOMWARE DETECTION EMPLOYING RULE-BASED ALGORITHMS

Hardi Sabah Talabani [a,*], Hezha M.Tareq Abdulhadi [b]

[a] Faculty of Medicals and Applied Sciences, Charmo University, Kurdistan Region, Iraq – hardi.talabani@charmouniversity.org
[b] Dept. of Information Technology, National Institute of Technology (NIT), Kurdistan Region, Iraq – Hezha.Abdulhadi@nit.edu.krd

Received: *4 Nov., 2021 /* ***Accepted:*** *12 Jan., 2022 /* ***Published:*** *16 Jan., 2022* ***https://doi.org/10.25271/sjuoz.2022.10.1.865***

ABSTRACT:

Cryptocurrencies have completely altered the digital transaction process all over the globe. Almost a decade after Satoshi Nakamoto generated the first Bitcoin block; many cryptocurrencies have been established. The Ransomware attack is a type of cybercrime and a class of malware that encrypts the files and prevents users from accessing their data or systems and demands payment for decrypting and retrieving access to their files. Ransomware data classification using present data mining and machine learning methods is difficult because predictions aren't always correct. We aim to build two models that effectively address these challenges and can diagnose and classify Ransomware attacks accurately, then compare the performance of the models. In this paper, we investigated the use of Rule-Based algorithms for mining Bitcoin Ransomware Data to classify Ransomware attacks in Bitcoin transactions. Employing Rule-Based techniques in detecting Bitcoin data is beneficial because the algorithms effectively classify non-linear datasets. The analysis was done on a Bitcoin dataset for 61,004 addresses selected from 29 Ransomware families and contained ten descriptive and decision attributes. Both Rule-Based algorithms were illustrated and compared on the dataset employing 10-fold cross-validation. Experimental results show that classification under partial decision tree (PART) algorithm performed better in different metrics than the Decision Table algorithm. It provides an accuracy of 96.01%, a recall of 96%, a precision of 95.9%, and an F-Measure of 95.6%. Experimental results propose that it is beneficial to further investigate the application of PART to predictive modelling tasks in Ransomware studies.

KEYWORDS: Bitcoin, Ransomware, Machine Learning, Rule-Based Algorithms, Decision Table, Partial Decision Tree (PART), Cybercrime.

1. INTRODUCTION

Satoshi Nakamoto announced Bitcoin as a cryptocurrency in 2008. The most known cryptocurrency is Bitcoin. By using an open-source code, Bitcoin was implemented in 2009. It is a digital banking system that doesn't have a physical central banking system and a particular country or radix. (Seow,2020) The most used and decentralized type of payment system is Bitcoin, whereas the public ledger is fully aided in a distributed way. A BlockChain is created by some unknown anonymous individuals that execute a protocol for maintaining and expanding a distributed public ledger that records Bitcoin transactions.

A BlockChain is a sequence of blocks that are implemented. Bitcoin transactions are entirely digital and anonymous to a significant extent (Seow,2020). Due to this predicament, many cybercriminals have turned to Bitcoin to conduct illegal operations such as Ransomware payments in a safe environment. (Mukesh,2018) Ransomware is a malignant program that attacks payment gateways in exchange for a ransom that must be paid.

Rule-based methods are a popular class of machine learning and data mining techniques. Their common objective is to discover patterns in data that are described in the form of an IF-THEN rule. We may distinguish between association rule discovery and predictive rule learning, depending on the sort of Rule that has to be discovered. (Mukesh,2018) A set of rules that collectively cover the instance space is generally desired in the latter case since they can predict every conceivable instance. To correctly predict the individuals or groups to whom Ransomware payments are made, Rule-Based techniques are used to go through previous transactions (Mukesh,2018).

In this paper, we investigate the use of two Rule-Based models: Decision Table and partial decision tree algorithm (PART), to classify the Bitcoin Ransomware, then compare the algorithm's performance according to the classification measurements criteria. Both algorithms performed reasonable outcomes in other studies and fields, so they will be tested on the Bitcoin dataset in this paper to detect the Ransomware and evaluate their performances. The remaining sections in this paper are organized as follows: Section 2 discusses the literature review. Section 3 explains the materials, methods, dataset, and algorithms used in this paper. Section 4 presents the experimental results, followed by a conclusion in section 5.

2. RELATED WORKS

Although Rule-Based algorithms have been widely used in the literature for various purposes and fields, they still lack development to be used in the Bitcoin Ransomware domain for classification problems. Also, different machine learning approaches were employed in identifying and detecting Ransomware. This section reviews those relevant papers that used machine learning techniques in Ransomware detection and Rule-Based models in other fields.

Akcora et al. (Akcora et al.,2020) employed topological data analysis methods for detecting recent malicious addresses in the Ransomware family. The authors generated a Bitcoin graph model in a directed weighted graph. Payments received to known Ransomware family addresses are used to identify new addresses belonging to the malware family. The Ransomware addresses are

* Corresponding author

initially grouped into 20,000 clusters. The resultant clusters are then examined to see any relationship between Ransomware families. To identify and predict Ransomware in money transactions, both Topological Data Analysis (TDA) and the DBSCAN clustering method (Density-Based Spatial Clustering of Applications with Noise) are utilized. For Ransomware transaction identification, their suggested technique significantly improved accuracy and recall compared to current heuristic-based methods and may be used to automate Ransomware detection.

Liao et al. (Liao et al., 2016) examined the analysis of the CryptoLocker malware family. A system that automatically identifies the ransom payments made to CryptoLocker Bitcoin addresses. BlockChain analysis and data fetched from online forums such as Reddit and BitcoinTalk were employed to analyse performance measurements on the data. The extraction of the timestamps based on the ransom payments by the victims is done. The time's ransom amount trends using this data were paid and analysed.

Huang et al. (Huang et al.,2018) conducted a measurement analysis of two-year data of Ransomware payments, including information about victims and operators. An extensive dataset was created using several data sources such as Ransomware files, victim telemetry, and a massive list of Bitcoin addresses. This information was utilized to Bitcoin-trail the victim from the time acquired the Bitcoins until the time the operators cashed out the Bitcoins. While comparing the two algorithms, the results of the existing one are improved and more accurate than the previous one.

Alhawi et al. (Alhawi et al.,2018) employed a novel approach for detecting Ransomware samples from features derived from network traffic communication; his approach was NetConverse which uses J48 based decision tree classifier. Their experimental results demonstrated that the proposed method detected better than other conventional machine learning algorithms as Bayes Network, K-Nearest Neighbor, Multi-layer perceptron, Random Forest, and Logistic Regression.

Kshirsagar et al. (Kshirsagar et al.,2019) used the two novel Rule-Based methods for intrusion detection. The feature selection used information gain with ranker attribute evaluator. The experimental results showed that PART and Decision Table performed well in detecting intrusion. At the same time, the Decision table performed outstandingly compared to the PART algorithms with 99.99% accuracy.

Hussein et al. (Hussein et al.,2021) worked on various machine learning and rule-based models to recognize fraud in credit card transactions. They tried to improve the performance of detecting fraud using multiple machine learning techniques by choosing the most appropriate algorithm for inclusion in the fraud identification system. In the experimental results, the two famous rule-based algorithms recorded a decent result with 81.95% accuracy for the Decision Table classifier and 81.37% for the PART classifier.

Sohail et al. (Sohail et al.,2019) applied Rule-Based classifiers (PART and Decision table) on a data mining platform to pinpoint potential diabetes and pre-diabetes in the initial medical monitoring via logistic regression prediction assessment analysis. Two hundred and eighty-one diabetes mellitus patients were assessed utilizing ten convenient and accessible non-invasive clinical characteristics gathered from four major hospitals in northern Nigeria. The results were published in the journal Diabetes Care. According to the experimental results, Rule-Based classifiers obtained maximum accuracy of 98.75%. Also, there was 0.98% for Error-Rate, 0.98% precision, 0.98% recall, and 98% an F-measure respectively.

Gaikwad and Thool (Gaikwad & Thool,2015) used the Bagging Ensemble technique in Intrusion Detection System implementation. Because of its simplicity, the Partial Decision Tree was employed as a starting point for classification. The relevant features are selected using an optimization algorithm to enhance the classifier's accuracy. Evaluation of the proposed intrusion detection system is done for classification accuracy, true positives, false positives, and the amount of time it takes to build a model. The proposed framework with the Partial Decision Tree (PART) algorithm to the other classifiers revealed that the proposed system with the Partial Decision Tree (PART) approach attained the maximum classification accuracy of 99.71 % employing cross-validation.

Alam et al. (Alam et al., 2021) examined ten various Machine Learning algorithms, including the Decision Table approach. The researchers tested the algorithm on the MovieLens dataset to evaluate the algorithms and determine the most appropriate one. Different evaluation metrics were employed after the classification process, including Kappa Statistic, F-measure, and Accuracy. Decision Table obtained a reasonable performance with 98.79% accuracy in the experimental result.

Previous studies deduce that various machine learning techniques have been used for Ransomware classification tasks. Some of those studies obtained outstanding results. For instance, (Alhawi et al.,2018) and (Huang et al.,2018) received good outcomes. On the other hand, other researchers have employed Rule-Based algorithms in various fields and obtained outstanding results. For instance (Sohail et al.,2019) (Gaikwad & Thool,2015) and (Alam et al., 2021). For this purpose, we decided to fill that gap and implement PART and Decision Table classifiers to identify and detect Ransomware in the selected dataset. Related results are mentioned in the following sections.

3. METHODOLOGY

In this section, the approaches, techniques, and dataset used in this paper have been demonstrated. Figure. 1 depicts the flow diagram of our method, and the following subsection explains each step of the investigation carried out in detail.

3.1 Bitcoin Dataset

The dataset used in this research on the Rule-Based algorithms was obtained from the Bitcoin transaction graph between January 2009 and December 2018. Collected daily network transactions and network linkages with fewer than 0.3 billion were filtered out since Ransomware sums were often more than this threshold. It comprises 61,004 addresses selected from 29 Ransomware families extracted from the UCI machine learning repository collected from January 2009 to December 2018 included extracting daily transactions from the network and constructing the Bitcoin over 24 hours (Akcora et al.,2021). The Bitcoin Heist Ransomware Address Dataset contains 10 descriptive attributes and decision attributes. The dataset summary is presented in Table 1.

Table.1 Description of the Bitcoin Ransomware Address dataset

Attr. ID	Attr. Name	Attr. Type	Attr. Description
1	Address	Text	Transactions Addresses. Tow Type of transactions in this dataset, Attack (Ransomware) or non-Attack (white)
2	Day	Number	The Year of transaction
3	Year	Number	The transaction day (1-365)
4	Length	Number	The total number of non-starter transactions on its longest chain.
5	Weight	Number	The sum of a fraction of coins that are constructed from the initial transaction and end up at the address.
6	Count	Number	The number of first-time transactions Associated with

			an address during a chain of events.
7	Looped	Number	The number of starting transactions associated with an address that has more than one immediate arc is shown.
8	Neighbors	Number	Some transactions contain the address as an output, whereas others do not.
9	Income	Number	The total amount of outputs from the coin to the destination address
10	Class/ Label	Text	The categories to which the transaction pertains are listed below. It is either a white or non-Ransomware category, which indicates that the transaction is risk-free or one of the 27 Ransomware categories, which indicates that the transaction is not risk-free or the subject of an assault.

Every transaction in the dataset has been gathered with a label pointing if the transaction is not attacking (white) or belongs to one of the 27 Ransomware families (attacks). In general, the used dataset is extremely imbalanced, and is considered a multi-class dataset. These completely different and extremely unbalanced labels are unequally distributed over the ten attributes indicated in Table 1, summarized in terms of how they are distributed, and the number of times each class is repeated in Table 2. The benign class (white) occupies the largest of the total dataset classes. The most frequent Ransomware family type is paduaCryptoWall among the other Ransomware families, which is repeated in lower percentages. This paper intends to study the comparisons of two Rule-Based classifiers (Decision Table and PART) in this completely imbalanced scenario. Captions:

Table.2 Frequency of occurrences of the class labels

Class	Label Name	Label Frequency
0	white	19591
1	paduaCryptoWall	12390
2	montrealCryptoLocker	9315
3	princetonCerber	9223
4	princetonLocky	6625
5	montrealCryptXXX	2419
6	montrealNoobCrypt	483
7	montrealDMALockerv3	354
8	montrealDMALocker	251
9	montrealSamSam	62
10	montrealGlobeImposter	55
11	montrealCryptoTorLocker2015	55
12	montrealGlobev3	34
13	montrealGlobe	32
14	montrealWannaCry	28
15	montrealRazy	13
16	montrealAPT	11
17	paduaKeRanger	10
18	montrealFlyper	9
19	montrealXTPLocker	8
20	montrealCryptConsole	7
21	montrealVenusLocker	7
22	montrealXLockerv5.0	7
23	montrealEDA2	6
24	montrealJigSaw	4
25	paduaJigsaw	2
26	montrealSam	1
27	montrealComradeCircle	1
28	montrealXLocker	1

3.2 Dataset Pre-processing

Before any experiments, it is better to check our data and prepare it for the tests. Since the used dataset contains many records of Bitcoin network transactions, it is possible to have a repetition or a duplicate of the same record, which may affect the performance of the algorithms used or reduce the accuracy. To get rid of this, we applied (RemoveDuplicates) filter on the dataset to remove duplicated and redundant data in Bitcoin network transactions of the same type.

3.3 Rule-Based Algorithms

Rule-based approaches are a prominent class of machine learning and data mining techniques. Their main objective is to discover regularities in data that can be represented as an IF-THEN rule (Fürnkranz et al., 2012). We may distinguish between association rule discovery and predictive rule learning based on the type of rule that has to be discovered. In the latter scenario, one is frequently interested in learning a set of rules that collectively cover the instance space to anticipate every potential occurrence (Lengyel,2015) (Qin,2009).

3.2. Decision Table

An ordered collection of IF-THEN rules, Decision Tables may be more compact and understandable than decision trees and are an accurate method for quantitative prediction. In comparison to the decision-tree-based method, the decision-table-based approach is more straightforward and requires fewer computer resources. A decision table is a basic assumption space that is easily understandable. Classifiers are generated by using decision tables. It evaluates feature subsets utilizing best-first search and cross-validation for assessment, and it reviews the whole feature set using best-first search (Kalmegh,2018).

3.3. Partial Decision Tree Algorithm (PART)

PART is a simplistic method, and it does not perform global optimization while developing proper rules. Using the divide and conquer strategy, it creates a rule by omitting the instances it encompasses and then continues to build recursive rules for instances rest until there are no more instances to be covered. The algorithm generates sets of rules known as *decision lists*, which are collections of rules organized in an ordered law. A current amount of information is compared to each direction in the list one by one, and the item is assigned to the category of the first matching rule (default is used if no control matches successfully). PART creates a partial C4.5 decision tree in each iteration and transforms the (best) branch into a rule. The PART classifier is an amalgamation of two algorithms: the RIPPER and the C4.5 rule learning (Mohamed et al.,2012).

3.4. 10-Fold Cross-Validation

The idea of cross-validation is as follows: it takes a specific data set and divides it into 10 separate parts for use in training and testing operations. The nine parts will be taken as training data, and the tenth (the final) part will be treated as the test data. This process continues until each of the ten parts is used as both training and testing data. Therefore, no data point is left in the used dataset until used nine times for training and once for testing (Talabani & AVCI,2018). In other words, cross-validation uses a periodical process to test and train itself on the rest of the data. Each of these ten parts is called a fold, meaning that each data set consists of 10 folds and ten sub-results. The final result is an

average of the sub-results (Talabani & AVCI, 2018). Completing the primitive division of the data in this form ensures that each partition or fold has obtained a correct percentage of the classes' values.

3.5. Evaluation Measurements

To measure the classification performances in machine learning, Accuracy, Precision, Sensitivity, and F-Measure are widely used metrics. The used metrics in this research are explained with their formulas below in detail (Geyik et al.,2021) (Varol & Abdulhadi,2018):

$$Accuracy = \frac{(TP+TN)}{(TP+FN+TN+FP)} \quad (1)$$

$$Precision = \frac{(TP)}{(TP+FP)} \quad (2)$$

$$Recall = \frac{(TP)}{(TP+FN)} \quad (3)$$

$$Sensitivity = \frac{(TP)}{(TP+FN)} \quad (4)$$

$$F-Measure = 2 * \frac{precision * sensivity}{precision + sensivity} \quad (5)$$

$$Error\ Rate = 1 - Accuracy \quad (6)$$

Where **True Positive (TP):** The number of ransom attacks that are correctly predicted as Ransomware attacks.
True Negative (TN): The number of non-Ransomware attack files correctly classified as non-Ransomware attacks.
False Positive (FP): The number of non-Ransomware attack files predicted as Ransomware attacks.
False Negative (FN): The number of Ransomware attacks predicted as non-Ransomware attacks.

In addition, we have shown other criteria in Tables 3, 4. Correctly classified instances mean the number of daily parameters were correctly predicted by the algorithms used; incorrectly classified instances represent the number of daily transactions that were incorrectly predicted by the algorithms used. Modelling time indicates the time for model building and classifying the dataset (Geyik et al.,2021) (Varol & Abdulhadi,2018).

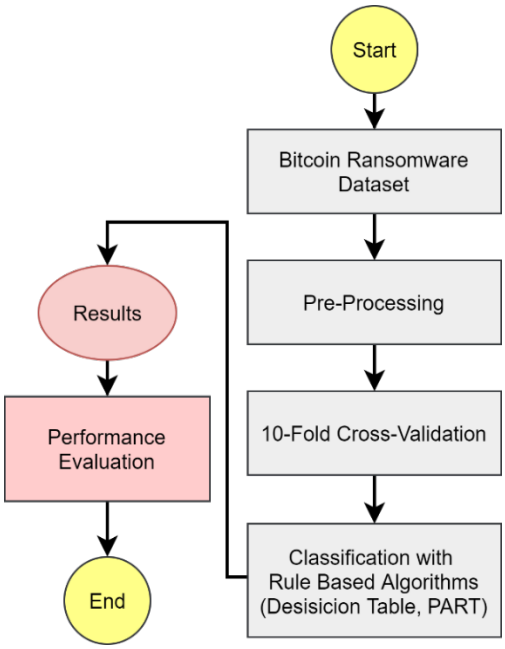

Figure 1. The Flowchart of the Proposed Work

4. EXPERIMENTAL RESULTS AND DISCUSSION

The fundamental objective of this study is to build a quick and accurate model for detecting Ransomware files in Bitcoin Dataset and specifying the most precise algorithm between Decision Table and PART algorithms. The classification process was done with a machine learning software created by the University of Waikato in New Zealand known as Weka (Waikato Environment for Knowledge Analysis). The GNU General Public License applies to Weka, is free to use. In addition to a set of visualization tools and algorithms for data analysis and predictive modelling, the Weka workbench also includes graphical user interfaces (GUIs) that allow for quick and simple access to this capability (Kotak & Modi,2020). We have implemented the work on Toshiba Tecra-Z40, a laptop Intel(R) Core (TM) i5-4300U CPU @ 1.90GHz 2.50 GHz, 8-GB RAM, Window 10 Professional 64-bit.

We have implemented the Rule-Based experiments conducted in this article on a huge dataset with many records. The Bitcoin Heist Ransomware dataset was first split using a 10-fold cross-validation method to divide the dataset into ten equal parts to implement and validate our proposed approach. Each part is used in both the training and testing process. This method aims to guarantee the randomness of the experiments and avoid any modelling issues with underfitting and overfitting. To evaluate the algorithm's performance after classification, Table 3 and Table 4 summarize each model's values of evaluation metrics.

Table 3. Time / Values of performance metrics of the algorithms

Rule-Based Algorithms	Decision Table	PART
Correctly Classified Instances	56718	58567
Incorrectly Classified Instances	4286	2437
Error Rate %	7.0258	3.9948
Modelling Time (Minutes)	1:43	26:49
Accuracy %	92.97	96.01

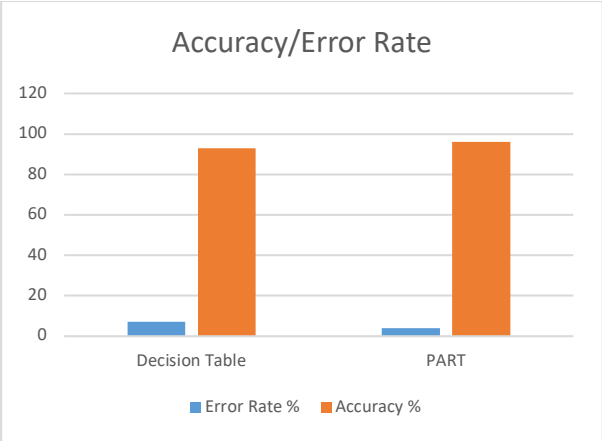

Figure 2. Error rate/ Accuracy of the algorithms

The experimental results in Table 3, Table 4, and Figure 2, Figure 3 show that the PART model has a better classification of the Bitcoin transactions dataset than the Decision Table classifier. We can see that the highest accuracy, precision, recall, and F-measure of 96.01%, 0.959, 0.960, and 0.956 respectively were obtained from the PART algorithm. And the correctly classified transactions out of the total transactions in the whole dataset for PART was 58567. We can also notice that the performance of the Decision Table obtained was lower than the PART algorithm, based on all evaluation metrics used, which was 92.97%, 0.924%, 0.930%, 0.925% for each accuracy, precision, recall, and F-Measure.

The correctly classified transaction out of the total transactions in the whole dataset for Decision Table was 56718. The PART algorithm obtained a 3.9948% error rate, whereas the error rate of Decision Table is higher than PART and it was 7.0258%. In addition, we can see a huge difference in model building and the classification time interval between the algorithms. Decision Table was faster than PART, which took 1:43 minutes. In contrast, the time needed for PART was 26:49 minutes.

Table 4. Evaluation Metrics for Validation Data

Rule-Based Algorithms	Decision Table	PART
Precision	0.924	0.959
Recall	0.930	0.960
F-Measure	0.925	0.956
TP Rate	0.930	0.960
FP Rate	0.012	0.007

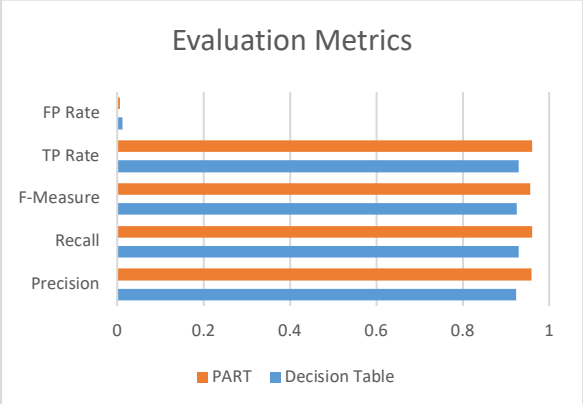

Figure 3. Performance Measurements of the algorithms

We can notice in Table.4 and Figure.3. The highest True Positive Rate, 0.960, was obtained from the PART model, whereas Decision Table obtained 0.930 in True Positive Rate. In contrast, the False Positive Rate of Decision Table was 0.012, and PART was 0.007.

Rule-Based models have been widely used for classification problems in machine learning and data mining. This paper presented two Rule-Based algorithms, PART, and Decision Table, to classify the Ransomware Bitcoin dataset. The Rule-Based models offer flexibility and can handle a big dataset as we used in this research. We have demonstrated that our classifiers can efficiently identify the Ransomware in a Bitcoin dataset with experimental studies and therefore obtained superior predictive performance. Although the classifiers performed well, a variation is still available among the classifier performances. Thus, the PART model got better results in classifying Ransomware than the Decision Table mode based on all evaluation metrics.

5. CONCLUSION AND FUTURE WORK

Experiments on Ransomware classification issues demonstrated the performance of the proposed models. From the results of our experiments, we conclude that the PART classifier had the highest performance for the Bitcoin Ransomware dataset showing an accuracy of higher than 96.01% compared to the Decision Table classifier with an accuracy of 92.97%. It is interesting to observe from Table 3 that PART has the highest evaluation performance. In all experiments, the Decision Table could not accurately predict all cases. At the same time, the PART model was able to classy almost all the Ransomware, as shown in Table 3. According to all the evaluation measures, the PART model's overall performance outperformed the Decision Table model. In this study, we limited ourselves to only using two different models. It would be more interesting to learn the impact of the choice of other models to investigate the Bitcoin Ransomware dataset classification further. We plan to employ a hybrid approach in detecting Ransomware attacks and extend our analysis for more and various versions of Ransomware attacks and real-time practice.

REFERENCES

Seow, K. T. (2020). Supervisory control of Blockchain Networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50(1), 159–171. https://doi.org/10.1109/tsmc.2019.2895345

S. D. Mukesh.(2018).An Analysis Technique to Detect Ransomware Threat. International Conference on Computer Communication and Informatics (ICCCI),pp1-5, doi: 10.1109/ICCCI.2018.8441502

Akcora, C. G., Li, Y., Gel, Y. R., & Kantarcioglu, M. (2020). BitcoinHeist: Topological Data Analysis for Ransomware prediction on the Bitcoin blockchain. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. https://doi.org/10.24963/ijcai.2020/612

Liao, K., Zhao, Z., Doupe, A., & Ahn, G. (2016). Behind closed doors: measurement and analysis of CryptoLocker ransoms in Bitcoin. 2016 APWG Symposium On Electronic Crime Research (Ecrime). doi: 10.1109/ecrime.2016.7487938

D. Y. Huang et al.(2018). Tracking Ransomware End-to-end.IEEE Symposium on Security and Privacy (SP).pp 618-631, doi: 10.1109/SP.2018.00047.

Alhawi, O. M., Baldwin, J., & Dehghantanha, A. (2018). Leveraging machine learning techniques for windows Ransomware network traffic detection in Cyber threat intelligence. Springer, Cham. pp. 93-106. doi: 10.1007/978-3-319-73951-9_5

Kshirsagar, D., & Shaikh, J. M. (2019, September). Intrusion Detection Using Rule-Based Machine Learning Algorithms. In 2019 5th International Conference On Computing, Communication,

Control And Automation (ICCUBEA) (pp. 1-4). IEEE. doi: 10.1109/ICCUBEA47591.2019.9128950

Hussein, N., Abbas, A., & Mahdi, B. (2021). Fraud Classification and Detection Model Using Different Machine Learning Algorithm. Tech-Knowledge Journal, 1(1).

Sohail, M. N., Jiadong, R., Muhammad, M. U., Chauhdary, S. T., Arshad, J., & Verghese, A. J. (2019). An accurate clinical implication assessment for diabetes mellitus prevalence based on a study from Nigeria. Processes, 7(5), 289. doi:10.3390/pr7050289

Gaikwad, D., & Thool, R. (2015). Intrusion Detection System Using Bagging with Partial Decision TreeBase Classifier. Procedia Computer Science, 49, 92-98. doi: 10.1016/j.procs.2015.04.231

Alam, M., Ubaid, S., Shakil, Sohail, S., Nadeem, M., Hussain, S., & Siddiqui, J. (2021). Comparative Analysis of Machine Learning based Filtering Techniques using MovieLens dataset. Procedia Computer Science, 194, 210-217. doi: 10.1016/j.procs.2021.10.075

UCI Machine Learning Repository: BitcoinHeistRansomwareAddressDataset Data Set. (2021). Retrieved 24 December 2021, from https://archive.ics.uci.edu/ml/datasets/BitcoinHeistRansomwareAddressDataset

Fürnkranz, J., Gamberger, D., & Lavrač, N. (2012). Foundations of rule learning. Springer Science & Business Media.

Lengyel, L. (2015). Validating rule-based algorithms. Acta Polytech. Hung, 12, 59-75.doi: 10.12700/aph.12.4.2015.4.4

Qin, B., Xia, Y., Prabhakar, S., & Tu, Y. (2009, March). A rule-based classification algorithm for uncertain data. In 2009 IEEE 25th international conference on data engineering (pp. 1633-1640). IEEE. doi: 10.1109/ICDE.2009.164

Kalmegh, S. R. (2018). Comparative analysis of the weka classifiers rules conjunctive rule & decision table on indian news dataset by using different test mode. International Journal of Engineering Science Invention (IJESI), 7(2Ver III), 2319-6734.

Mohamed, W. N. H. W., Salleh, M. N. M., & Omar, A. H. (2012, November). A comparative study of reduced error pruning method in decision tree algorithms. In 2012 IEEE International conference on control system, computing and engineering (pp. 392-397). IEEE. doi: 10.1109/ICCSCE.2012.6487177

Talabani, H., & Engin, A. V. C. I. (2018, September). Performance comparison of SVM kernel types on child autism disease database. In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP) (pp. 1-5). IEEE. doi: 10.1109/IDAP.2018.8620924

TALABANI, H., & Engin, A. V. C. I. (2018, September). Impact of various kernels on support vector machine classification performance for treating wart disease. In 2018 International Conference on Artificial Intelligence and Data Processing (IDAP) (pp. 1-6). IEEE. doi: 10.1109/IDAP.2018.8620876

Geyik, B., Erensoy, K., & Kocyigit, E. (2021, January). Detection of Phishing Websites from URLs by using Classification Techniques on WEKA. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 120-125). IEEE. doi: 10.1109/ICICT50816.2021.9358642

Varol, C., & Abdulhadi, H. M. T. (2018, December). Comparision of string matching algorithms on spam email detection. In 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT) (pp. 6-11). IEEE. doi: 10.1109/IBIGDELFT.2018.8625317

Kotak, P., & Modi, H. (2020, October). Enhancing the Data Mining Tool WEKA. In 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1-6). IEEE. doi: 10.1109/ICCCS49678.2020.9276870