Bilad Alrafidain Journal for Engineering Science and Technology
https://dx.doi.org/10.56990/bajest/2024.030202
ISSN: 2073-9524
Page 17-25

Strengthen Method Attacks Detection in Network using VFDT Algorithm

Naseer Alwan Hussein*
Computer of Engineering Networks, College of Fine Arts, University of Diyala, Iraq
am.naseeralwan@uodiyala.edu.iq

Abstract

With the tremendous progress in organizing data over networks, network security has become more important than ever in modern technologies. Hacking represents a serious security risk in data transmission and reception. Sometimes there are new types of intrusions that are constantly evolving and detection systems have difficulty identifying them. In this research, we proposed an organization-based Abnormality Discovery Framework (OADF) system based on training the Very Fast Decision Tree (VFDT) algorithm to detect attack detection (AD). Exploratory outcomes on the KDD cup'99 data mining list show that our system is profoundly material to arrange interruption acknowledgment. It has been demonstrated that the proposed methodology performs better in terms of acknowledgment rate, deceiving issue rate, and computational time compared with various techniques. In this study, another type of previously prepared test group has been used, SSENet-2011, and we trained these two types separately to reach optimal training. The purpose of this work is to discover the largest number of attacks or penetrations that occur on the internet. An exceptional, very fast decision tree (VFDT) algorithm was utilized to build the classifier for attacks. Exceptionally, the VFDT algorithm is one of the data mining algorithms that arrangement with high data streams in an extremely brief time frame. Trial results exhibited that the proposed Organization based Abnormality Discovery Framework (OADF) system is exceptionally effective in detecting known and obscure assaults by 93%.

Keywords: Very Fast Decision Tree Algorithm (VFDT), classifier mode, network security, attacks detection (AD), kddcup'99 dataset, and SSENet-2011 dataset.

Article history: Received: 21-4-2024***, Accepted:***23-5-2024***, Published:*** 15-9-2024

1. Introduction

Information is becoming more and more readily available for immediate access and natural handling among our people as a whole. More information is being stored on computers as a result of these solicitations. The problem of unauthorized access and data manipulation has decreased with the spread of sensible laptops and PC networks. Due to their extensive understanding of how structures function, intruders have developed a rare talent for identifying flaws in systems and taking advantage of them to obtain such. The fact that they can do anything on the structure is a weakness. They also frequently employ strategies that are challenging to understand and implement. Because of the intruder's displays, PC structures are therefore unlikely to be protected in the near future. As a result, we need to have mechanisms in place to detect security breaches, that is, identify intruders and their methods of interfering. An Assaults Location Framework (Promotions) examines the procedures inside a framework for dubious methods of behavior or schemes that might indicate abuse or assault of the system. Hack disclosure techniques fall into two main categories: attack detection (AD) and abuse acknowledgment. The earlier one evaluates the collected data and compares it to a model that has been depicted. It can determine what is considered "normal" work lead find a clever way to perceive network desires. Soon after that is dim Abuse: Differentiating evidence depends on the markings made for references in terms of attacks, it is often only as remarkable as the attack's database indicates that it is used for relationships misuse of identifiable proof can't discriminate

* Corresponding author: Am.naseeralwan@uodiyala.edu.iq

ISSN: 2073-9524
Page 17-25

between new attacks despite having a low false issue rate. Periodicity disclosure, however, is capable of identifying dark attacks and has a somewhat greater deluding problem rate. We revisit the introduction of classifiers in this study. When prepared to recognize indications of clear-cut attacks. The attacks that have been replicated are listed below. Abstaining from Administration Assault (FAA): this type of assault involves memory or coping with issues excessively engaged or excessively full to ignore real requests or denials genuine customers agree to a machine.

Testing Assault (TA): this is an effort to assemble data about an association of Computers are the indisputable means of getting around its security measures. Client to Root Assault (C2R): this type of exploit starts with the attacker gaining access to a typical client's record on the system (which may be obtained by social engineering, word reference attack, or password sniffing) and may use a flaw to bind the design's underlying consent. When an attacker can transmit packets to a machine over a network but does not have a record on that system, they can exploit a vulnerability to gain local access as a client of that machine, which is known as remote to nearby assault (R2N) [1], [2] and [3].

An overtaking identifier framework called Review Information Examination and Data Mining (EDAM) is used. To identify overwhelming using information mining techniques. It keeps preparatory data that has been known to be freed from attacks at first. Next, it groups attacks, ambiguous behavior, and deceptive issues using an algorithm. Specifically, (EDAM) has a few useful capabilities;

- Characterize a thing as a known assault.
- Characterize a thing as a typical occasion.
- Characterize a thing as an obscure assault.

Research work gives a close examination of a few inconsistencies in discovery plans for recognizing novel net attack location distinguishing pieces of proof. We present exploratory outcomes on data Revelation Information Mining KDD Cup'99 data index. The exploratory outcomes have shown that utilizing the Very Fast Decision Tree Algorithm VFDT model is significantly more productive in the identification of attacks, contrasted, and other grouping methods, as well as utilizing SSENet-2011. Data combination to lead new analyses to find ways of stronger the control of assaults Below we review the VFDT algorithm that was worked on using KDD cup'99 data and SSENet-2011 data within the proposed system that it was built in our research paper. The rest research paper is coordinated as follows. The work area will be the proposed Organization based Abnormality Discovery Framework (OADF), and give general data about attacks. The third presents the trial results.

The fourth segment closes the research paper. We will examine the speedy decision tree VFDT Algorithm, which is one of the specific Algorithms in assessment, and out of results with high accomplishment rates stood out from Algorithms that Work in a competitive field especially speedy. The decision Tree algorithm is perceived by its ability to learn through planning using data KDD cup99 dataset and SSENet-2011 through which the estimation can work and show extraordinary outcomes. However, the computation was picked for this work to distinguish attacks online to control and confine malignant attacks that pollute laptops, as it translates the attacks into a couple of levels, including outstandingly unsafe. Ones that can’t be controlled, fast processor needs, and sensitivity. dangerous ones that can be overseen suitably, has the preferred limit over separate data in a very short period, which helped in its utilization in a few projects by mining and sifting information and creating great outcomes for the work. The extremely Very Fast Decision Tree Algorithm VFDT was utilized and refreshed using two kinds of preparing information and this component in which there are vacant cells. The leaves of a tree develop with expanding preparation of the calculation in a simple and basic manner. The following is a chart showing the activity of the extremely quick choice tree calculation. It works in two frameworks, the first without an organization association and the second with an organization association [4]. The implementation consists of two phases, the first presents the training stage and the second presents the testing stage as displayed in Figs. 1 and 2.

ISSN: 2073-9524
Page 17-25

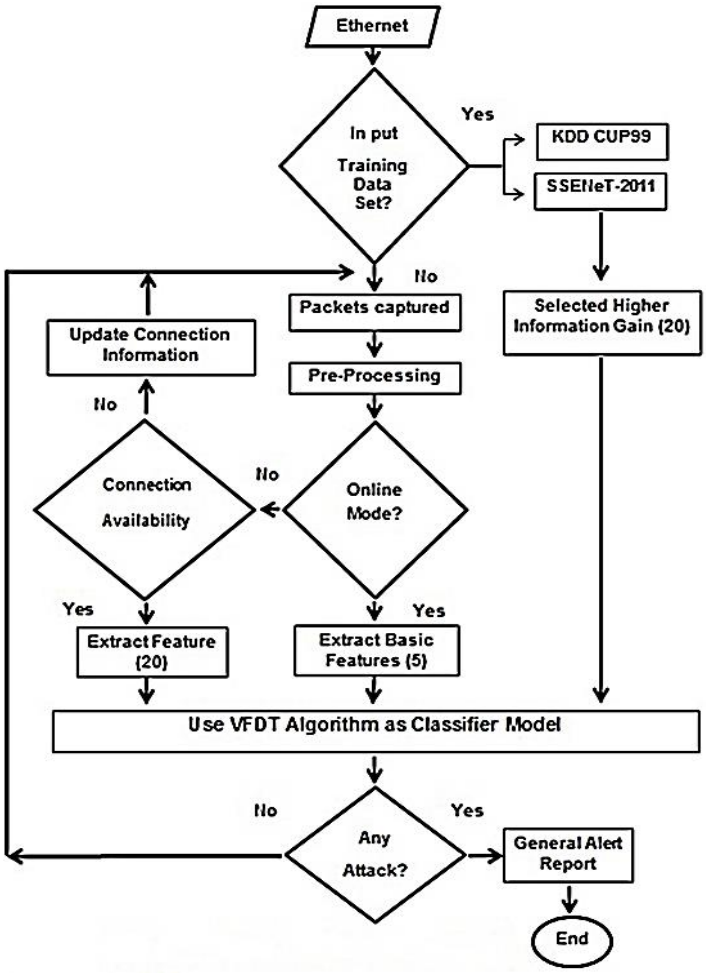

Fig. 1 Block Diagram of the Training Phase

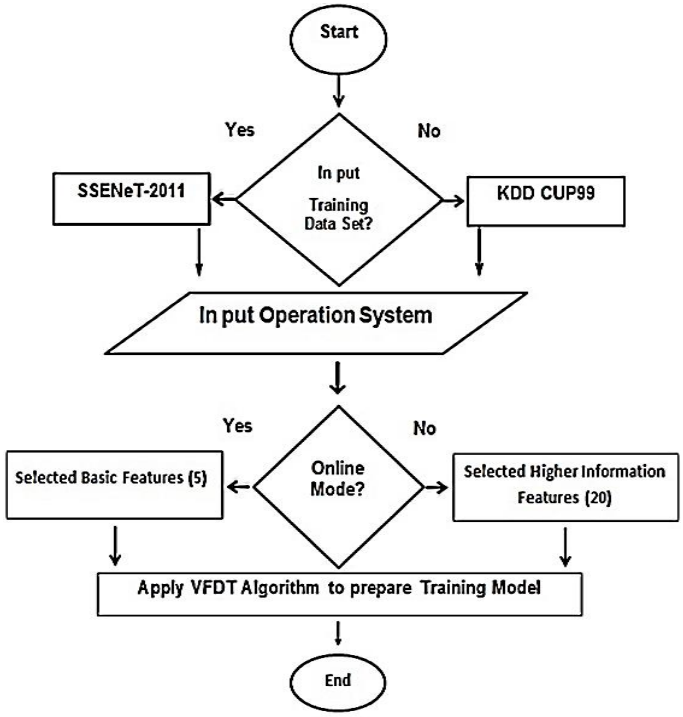

Fig. 2 Block Diagram of the Testing Phase

ISSN: 2073-9524
Page 17-25

2. Materials and Methods

There is a possibility that the suggested OADF framework will function in separate modes or on the Internet. This depends on the kinds of components that are used to test and prepare the framework. In a sense, the online mode handles the essential components of an extraction by using the header of the gatherings. (Display type, organization, flag, src_. bytes, dst_bytes, for instance.). These elements are made possible by affiliation records that are incorporated into the model. It is imperative to select the association's option as soon as it becomes available and to respond appropriately. The detached mode handles the fundamentally measurable elements removed from the package content and header. When the meeting is over, the quantitative elements are eliminated and they are constructed as affiliation records. To test their courses, endless packs are meant to recognize precise connections and then preserve them in the archive as affiliation records that are placed in the depicted model as typical.

This mode provides precise outcomes, yet due to the obstacle of time, sufficient affiliation information is needed for both modalities to establish the suggested model. Following that, the system updates and gathers information at any given time to establish a decision tree that may be used to gather attacks in a coordinated manner. VFDT is a popular show that uses Hoeffding tree-based data mining Different request learning processes have been. The decision tree learning methodology was suggested and is frequently utilized since it is quick and efficient. the representation of classifiers that it determines is genuinely seen.

This data stream expands dynamically as new information arises and is presented; VFDT permits the use of either information gain or the addition of new data to the list as a trademark appraisal measure. It combines several computation enhancements into one. Since it can reliably produce without holding it together for the existence of the overall big number of models, VFDT does not assemble models in such a mindset. Fig. 3 shows the suggested Organization-based Abnormality Discovery Framework (OADF).

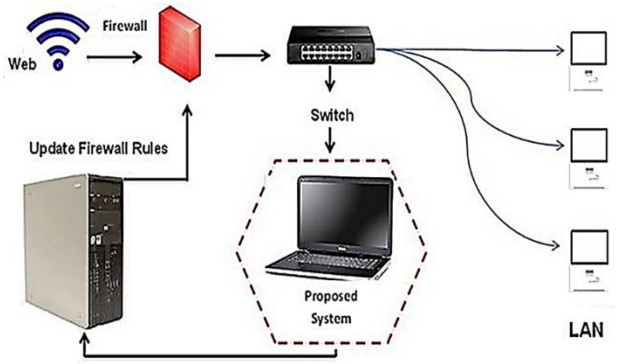

Fig. 3 The design of the proposed Organization based Abnormality Discovery Framework (OADF)

3. Results and Discussion

In this segment, we sum up exploratory outcomes in distinguishing network interruption utilizing the VFDT algorithm over the KDD Cup'99 dataset and SSENet-2011 data set, first portray the informational index utilized in quite a while and afterward examine the outcomes last, then, compare the obtained results of KDD Cup'99 dataset collection with the outcomes of SSENet-2011 dataset acquired. We completed four trials on the KDD cup'99 dataset and SSENet-2011 dataset to obtain the best outcomes. The conditions of these tests are according to the accompanying:

- Investigate 1: We used the KDD cup'99 dataset and the SSENet-2011 dataset for getting ready, but we picked 20 features specifically.
- Investigate 2: We utilized the KDD cup'99 dataset and the SSENet-2011 dataset for preparation, yet we picked 15 highlights.
- Investigate 3: We utilized 10% of the KDD Cup'99 dataset or the entire SSENet-2011 dataset for preparation, and we chose 15 or 20 elements.

ISSN: 2073-9524
Page 17-25

- Try 4: We utilized the KDD cup'99 dataset and the entire SSENet-2011 dataset for preparation, yet we picked just the essential highlights. The confusion matrix (CM) showed used to rate of detection attacks of AD and it has the accompanying entries:

Table 1: Confusion Matrix (CM)

Confusion Matrix (CM)		
	Normal	Intrusion (Attacks)
Normal	TN	FP
Intrusion (Attacks)	FN	TP

- True Positive (TP): Number of affiliations that were precisely named attacks.
- True Negative (TN): Number of affiliations that were precisely named common.
- False Positive (FP): Number of normal affiliations that were designated attacks.
- False Negative (FN): Number of attack affiliations that were appointed regularly.

The degree of the really accumulated attacks is represented by the DR, and it is calculated by equation 1. The degree of an incorrectly requested relationship, as demonstrated in, is the FAR that can be calculated by equation 2.

$$DR = \frac{tp}{tp + fn} \quad (1)$$

and,

$$FAR = \frac{fp}{fp + tn} \quad (2)$$

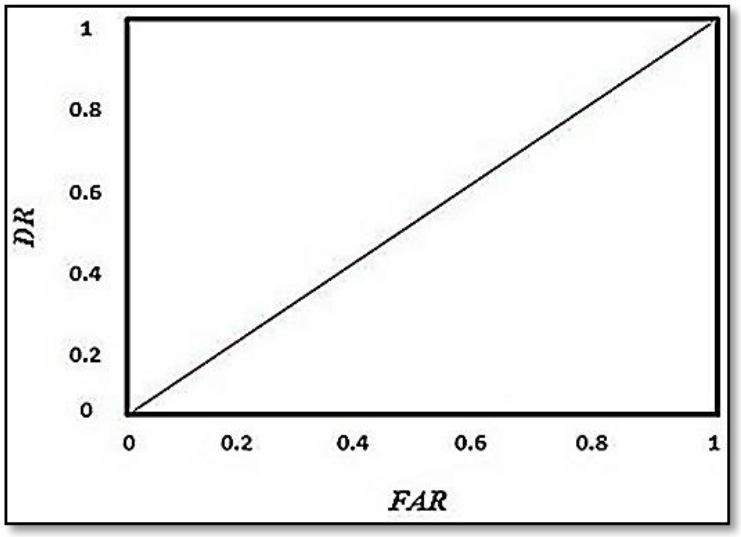

Fig. 4 Receiver operating characteristic (ROC)

The receiver operating characteristic is also shown in Fig. 4, which shows that the figure is divided into two halves. The upper half represents the percentage of correct attack detections, while the lower half represents the system's failed detections.

The Veritable Positive Rate (TP%) for the attacks FAA, test, C2R, and R2N is defined as the total number of connected assault TPs divided by the total number of TPs. The results of these tests should be apparent in Table 2.

Table.2 The results of these experiments

No.	**Results**	**DR %**	**FAR %**	**FAA TP %**
1	TA TP %	93.83	0.608	93.10
2	C2R TP %	79.04	22.84	36.64
3	R2N TP %	92.56	0.720	88.10

The main initial result demonstrated that the suggested OADF system. Utilizing the VFDT estimation, a DR percent of 93.83% for attacks was obtained; the cultivated FAR rate is 0.608%. Furthermore, the TP% for FAA, Test, TA, and C2R are, respectively, 93.10%, 79.04%, 22.84%, and

ISSN: 2073-9524
Page 17-25

36.64%. The postponed assessment's consequence was near when the Chief investigated it, but R2N's acknowledgment rate dropped to zero. According to the third exploratory result, the suggested OADF system achieved a FAR percent of 0.993% and a DR percent of 95.53 percent additionally, the TP%. For the C2R, TA, FAA, and Test 81.6%, 24.2%, 75.6%, and 96.4%, respectively. That was demonstrated by the fourth exploratory finding.

The Veritable Positive Rate (TP%) for the attacks FAA, test, TA, and C2R is defined as the total number of connected assault TPs divided by the total number of TPs. The results of these tests should be apparent in Table 3.

Table: 3 Execution measurements for the four trials

Experiments	DR%	FAR	FAA TP%	TA TP%	C2R TP%	R2N TP%
Expirmint1	93.825	0.608	93.106	79.04	22.84	36.64
Expirmint2	92.56	0.720	88.1	80	0	33.6
Expirmint3	95.532	0.993	96.4	24.2	75.6	81.6
Expirmint4	90.34	0.78	92.67	75.4	0	31.32

To the extent that structure precision, gathering speed, and memory segment are concerned, going with restrictions is examined to make sense of our outcomes exactly. Recipient Working Brand name (ROC) is a technique derived from a quantified choice hypothesis about the disclosure of electronic signs. It makes use of examining the classifiers' farthest reaches. Plots of ROCs are made of Sorts that are dispersed according to the DR and FAR depiction speeds. The suggested OADF structure executes with a high degree of accuracy when used in distinct systems. In Fig. 5, the structure's ROC bend is shown.

The objective is to perceive many assaults while limiting the FAR hour Figure 5 makes it quite clear that our system is far more capable of seeing. The majority of the attacks had a low FAR speed of 0.608% and a DR speed of 93. 83%. Fig. 5 displays the ROC twist for the OADF structure, while Fig. 6 represents the time it takes to sort things out compared to the quantity of affiliations. The OAD.F architecture uses 100,000 relationships at 4.09 seconds, but 100,000 connections at 38.97 seconds.

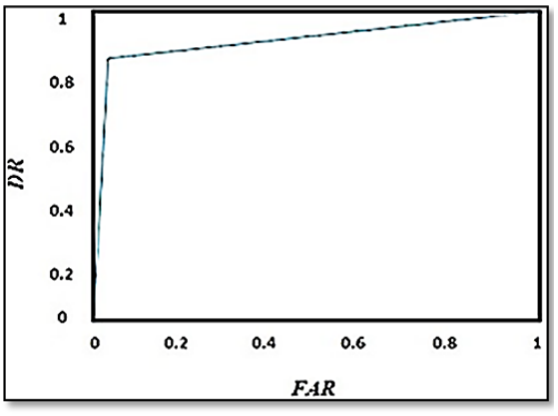

Fig. 5 ROC Curve for OAVD System

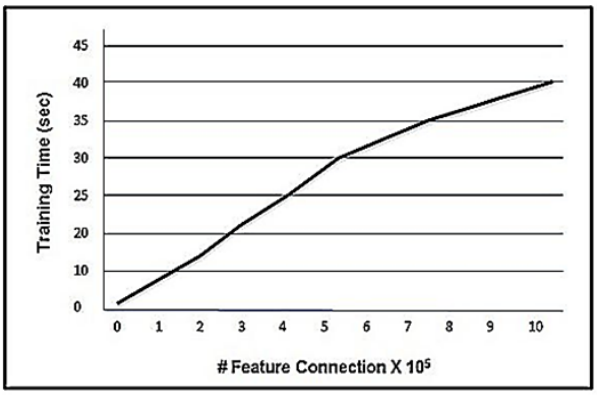

Fig. 6 Training Time versus number of affiliations.

Furthermore, Fig. 7 illustrates the number of nodes that are made through the use of the VFDT algorithm. The confusing stretch at the beginning of the chart marks the beginning of a tree's improvement, to be able to describe a large number of attacks in a similar tree; a sufficient number of associations must be accumulated. It happens gradually, one tiny step at a time, and the amount of concentration is essentially constant.

ISSN: 2073-9524
Page 17-25

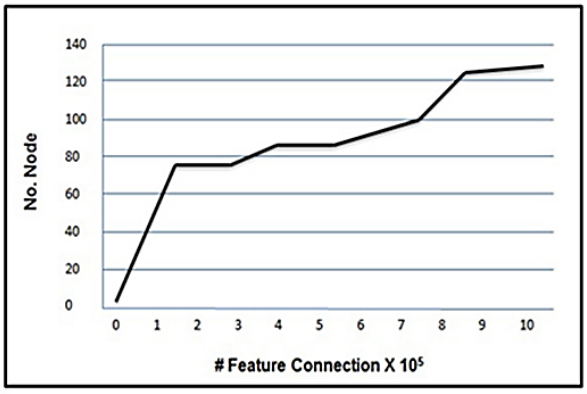

Fig. 7 Number of Nodes versus the number of affiliations

Table: 4 A comparison of the proposed system with another classification system that used the KDD Cup '99 dataset and SSENet-2011 dataset.

No.	System	Training Datasets size	Testing Datasets size	DR %	Training Time (TT) in sec.	Training Time Per Example
1	OADF	1.074.985	67.688	93.825	39.88	0.000003
2	GP	24.780	311.028	98	6480	0.2615
3	C4.5	49.596	15.437	92.06	15.85	0.0003
4	FL	54.226	56.226	91.25	87.9	0.0016
5	PART	444.458	49.384	46.67	48.8	0.0001
6	K-MEANS	55.000	25.000	86	13	0.0002
7	J48	49.596	15.437	92.06	15.85	0.0003
8	APRIORI	444.458	49.384	87.5	18.94	0.00004
9	BIN	49.596	15.437	90.62	6.28	0.0001
10	LBK	49.596	15.437	92.22	10.63	0.0002
5	PART	444.458	49.384	46.67	48.8	0.0001
6	K-MEANS	55.000	25.000	86	13	0.0002

4. Conclusion

The suggested Organization based Abnormality Discovery Framework (OADF) in this survey helped to manage devious attacks. There are two ways to apply this philosophy: online and offline. The suggested OADF system classifies the relationship as regular or assault and makes use of the very fast decision tree algorithm VFDT. In fact, the suggested OADF system has perfect time and outperforms other structures in terms of arranging time per model merits concentrating on it. The proposed system offered the option to view a massive portion of the attacks at a low FAR speed of 0.608% and a DR speed of 93.83%. We used two data kinds during the training stage: the SSENet-2011 dataset and the KDD Cup'99 dataset.

Conflict of interest

The author declares that the publishing of this article does not include any conflicts of interest. Furthermore, the author has strictly adhered to ethical problems such as plagiarism, informed consent, misconduct, data fabrication and falsification, multiple publishing and submission, and redundancy.

References

[1]. DONG, Zhenjiang, et al. Random forest based very fast decision tree algorithm for data stream. ZTE Communications, 2020, 15.S2: 52-57.

[2]. JIA, Shuangying. A VFDT algorithm optimization and application thereof in data stream classification. In: Journal of Physics: Conference Series. IOP Publishing, 2020. p. 012027.

ISSN: 2073-9524
Page 17-25

[3]. KHINE, Aye Aye; KHIN, Hint Wint. Credit card fraud detection using online boosting with extremely fast decision tree. In: 2020 IEEE conference on computer applications (ICCA). IEEE, 2020. P. 1-4.

[4].YACOUB, MOSTAFA M.; REZK, AMIRA; SENOUSY, M. Adaptive classification in data stream mining. Journal of Theoretical and Applied Information Technology, 2020, 98.13: 2637-2645.

[5]. Dhieb, N., Ghazzai, H., Besbes, H., & Massoud, Y. (2020). A secure ai-driven architecture for automated insurance systems: Fraud detection and risk measurement. IEEE Access, 8, 58546-58558.

[6]. Soe, Y. N., Feng, Y., Santosa, P. I., Hartanto, R., & Sakurai, K. (2020). Towards a lightweight detection system for cyber-attacks in the IoT environment using corresponding features. Electronics, 9(1), 144.

[7]. GHIASI, Mohammad M.; ZENDEHBOUDI, Sohrab; MOHSENIPOUR, Ali Asghar. Decision tree-based diagnosis of coronary artery disease: CART model. Computer methods and programs in biomedicine, 2020, 192: 105400.

[8]. Koulinas, G., Paraschos, P., & Koulouriotis, D. (2020). A Decision Trees-based knowledge mining approach for controlling a complex production system. Procedia Manufacturing, 51, 1439-1445.

[9]. Moayedi, H., Jamali, A., Gibril, M. B. A., Kok Foong, L., & Bahiraei, M. (2020). Evaluation of tree-base data mining algorithms in land used/land cover mapping in a semi-arid environment through Landsat 8 OLI image; Shiraz, Iran. Geomatics, Natural Hazards and Risk, 11(1), 724-741.

[10].Sahmoud, S., & Topcuoglu, H. R. (2020). A general framework based on dynamic multi-objective evolutionary algorithms for handling feature drifts on data streams. Future Generation Computer Systems, 102, 42-52.

[11]. GARCÍA-MARTÍN, Eva, et al. Energy-aware very fast decision tree. International Journal of Data Science and Analytics, 2021, 11: 105-126.

[12]. Ducange, P., Marcelloni, F., & Pecori, R. (2021). Fuzzy Hoeffding decision tree for data stream classification. Int. J. Comput. Intell. Syst., 14(1), 946-964.

[13]. LEFA, Mai; HATEM, A. B. D.; SALEM, Rashed. Enhancement of Very Fast Decision Tree for Data Stream Mining. Studies in Informatics and Control, 2022, 31.2: 49-60.

[14]. BA'ABBAD, Ibrahim; BATARFI, Omar. Proactive Ransomware Detection Using Extremely Fast Decision Tree (EFDT) Algorithm: A Case Study. Computers, 2023, 12.6: 121.

[15]. CHEN, Bin; CHEN, Qingming; YE, Peishan. Information-based massive data retrieval method based on distributed decision tree algorithm. International Journal of Modeling, Simulation, and Scientific Computing, 2023, 14.01: 2243002.