JSCDM

http://penerbit.uthm.edu.my/ojs/index.php/jscdm

e-ISSN : 2716-621X

Journal of Soft Computing and Data Mining

Feature Selection of Distributed Denial of Service (DDos) IoT Bot Attack Detection Using Machine Learning Techniques

Sharifah Shahmim Syed Othman[1], Cik Feresa Mohd Foozy[1*], Siti Noor Baini Mustafa[2]

[1]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, Johor, MALAYSIA

[2]Book Hack Enterprise,
E-02-01, Bayu Residence 1, Bandar Baru Nilai,71800 Nilai, Negeri Sembilan,MALAYSIA

DOI: https://doi.org/10.30880/jscdm.2023.04.01.006
Received 16 March 2023; Accepted 20 April 2023; Available online 25 May 2023

Abstract: Distributed Denial of Service (DDoS) attacks can be made through numerous mediums, becoming one of the biggest threats to computer security. One of the most effective approaches is to develop an algorithm using Machine Learning (ML). However, the low accuracy of DDoS is because of the feature selection classifier and time-consuming detection. This research focuses on the feature selection of DDoS IoT bot attack detection using ML techniques. Two datasets from NetFlow, NF_ToN_IoT and NF_BoT_IoT, are manipulated with two attributes selection: Information Gain and Gain Ratio, and ranked using the Ranker algorithm. These datasets are then tested using four different algorithms, such as Naïve Bayes (NB). K-Nearest Neighbor (KNN), Decision Table (DT), and Random Forest (RF). The results were compared using confusion matrix evaluation Accuracy, True Positive, True Negative, Precision, and Recall. The result from two datasets is selected by the Top 4, Top 8, and Top 12 feature selection. The best overall classifier is Naïve Bayes, with an accuracy of 97.506% and 90.67% for both datasets NF_ToN_IoT and NF_BoT_IoT.

Keywords: Machine Learning, DDoS, feature selection, Information Gain, Gain Ratio, Naïve Bayes, KNN, Decision Table, Random Forest

1. Introduction

The usage of technology is introducing new difficulties and dangers to cybersecurity. Organization has to deal with higher potential of cyberattacks which is usual for hackers, attackers and fraudsters to take advantage of situations when people are defenseless [1]. There was 220 percent rise in spam email and 260% increase in harmful URLs and United States shows the highest number of attacks and the most attacks are DDoS attacks. Hackers flood organization' systems or websites with false or bot users to disrupts the operation and communication channel. IoT attacks was ranked fourth highest cyberattacks in 2021 and most of them happen in government and healthcare organization' [2]. There are three main categories of DDoS attacks which are the application layer attacks, protocol attacks and volumetric attacks [3]. Variety of features and datasets has been made by past researchers and experiments, but all these features require high computing power and time consuming. The availability of labelled network traffic dataset in identifying threats through incoming network data requires large amount of network traffic data which lead to low detection precision and high false positive rates. Therefore, this research focusses on feature selection manipulation in order to improve the low accuracy of DDoS detection using ML techniques.

The objectives of this research are (1) to study the feature selection techniques to improve the parameter evaluation detection, (2) to select significant features of DDoS IoT bot attack using features ranking algorithm information gain and gain ratio, (3) to test and validate the features using confusion matrix accuracy, true positive true negative, precision and

*Corresponding author: feresa@uthm.edu.my

recall. This study will contribute in two areas which are the best classifier to increase detection time using NetFlow dataset, and the manipulation of significant features to increase detection accuracy.

This article is arranged as follows. Section 2 reviews related works from previous researchers. Section 3 explain the research methodology. Section 4 discuss the experimental results, discussion and analysis throughout this research. Lastly, Section 5 conclude the research and suggest future works possible to expand this research.

2. Related Works

An overview of current methods for DDoS feature selection attack detection is provided in this section.

2.1 DDoS

DDoS is one of the deadliest cyberattacks from cybercriminals targeted numerous computer system with flood request or false or bot request in massive amount of traffic at a time. The target attacks often happen in server, website, social medias or other network resources in order to disrupt traffic flow and create a denial of service to users of the targeted resource [4]. This causes a site to slow to a crawl or even crash which preventing the legitimate traffic from reaching the site. This kind of attempts can do serious damage to the business or organization. DDoS attacks are most dangerous when they target essential national infrastructure such as electricity, water supplies, transportation networks and healthcare organizations [5]. These intrusions can be done through wide range of interests varying from dissatisfaction and hacktivism to major financial damage. Differences between hacking and DDoS attacks are their purpose of attacks, where several types of malwares such as ransomware and scareware are done to steal money from victims or organizations, DDoS attack is done to cause confusion, chaos and disruption towards the organizations system. The amount of downtime damage they may inflict is one of the biggest reason for DDoS became the main topic of discussion in tech forum and websites [6].

2.2 Types of DDoS

Even though DDoS attack target are usually to overworking the system, there are several methods to do it. Three main categories of DDoS are application layer attacks, protocol attacks and volumetric attacks. Application layer attacks happen when different bots repeatedly request the same resources from server at the same time making the system overwork causing downtime and crash of server, such as HTTP flood using different IP address. Protocol attacks mostly happen around server's resources such as load balancer, routing engines or firewalls. Example of protocol attacks are SYN flood where the server is flooded with multiple SYN packets with forged IP addresses. Volumetric attacks happen when the server is flooded with massive traffic and fill up the available bandwidth, such as DNS amplification where attackers use fake IP address to send queries to DNS server causing chaos on the target server for DNS replies.

2.3 DDoS Attacks

DDoS attack affect online services by making the service unavailable to the users. DDoS detection is expensive to maintain, yet it is the most effective method to prevent vulnerable network security systems. The system presents a huge explosion in terms of the impact that will deny the user access, by overloading and signal to a total non-operational state. DDoS attack can be detected by examining or monitoring the normal network traffic flows or conditions. The existing methods have limited resources such as large dataset, low accuracy of the used algorithm, hardly or never updated software and supervised learning. Traditional network-centered security has relied on predefined signature or system representation for known threats [3], [7]. Researchers and organizations contribute their time to find ways to improve the defensive layer to fight these attacks. Machine learning (ML) is a model consisting of several dataset and models to represent a simulation of an attack. ML is used to train the models to detect attacks and help to predict the actual impact of the real-life threat and use the analysis to prevent or reduce the damage. In this paper, machine learning techniques are reviewed in order to list out methods used in ML to detect DDoS attacks [8]. The usual DDoS symptoms are large amount of traffic coming from clients, with mostly same characteristic such as browser type, IP range, device port and location. Another common symptom is the server repeatedly crashing for no reason and the website took too much amount of time to respond request.

2.4 IoT

The Internet of Things (IoT) is a sophisticated automation and analytics system that uses big data, artificial intelligence, networking, sensing technologies for integrated system for products and services. IoT systems has exceptional versatility and the ability to sustain and function in variety of sectors in whole wide world especially technology related industries. Smart gadgets helps them to improve data collection, operations and automation across the Internet and lifestyles [9], [10]. IoT solutions systems allow users to gain greater analysis, automation and integration in order to increase precision and range of these field. IoT increase the amount of usage of networking and robotics technologies with artificial intelligence and active engagement. AI enhances every characteristic of IoT with data

collection, AI algorithm and networks to assist daily routines at home or workplaces. Various benefits of IoT in lifestyle and businesses offers improved customer engagement, technology optimization, reduced waste to more effective management resources and enhanced data collection. Some drawbacks of IoT are the issues in security, privacy, complexity, flexibility and compliance seems incredibly challenging when defining the standards and usage policies which exposes users to various kinds of attackers and software breach [11].

2.5 Related Works On ML Techniques

There are many ML techniques and algorithm classifier and method has been applied to increase DDoS attack detection. Akanji [12] uses Genetic algorithm and SVM and NetFlow dataset to detect a slow HTTP DDoS attack. Awan [13] use Random Forest, MLP and Scikit with SparkML dataset to get the DDoS attack detection in real-time. Swe [14] use Random Forest, KNN, MLP and PART algorithm with NetFlow datasets to detect DDoS defensive mechanism reaction. Wang [15] use Dynamic MLP, SBP-MLP algorithm with NSL-KDD dataset to improve availability of modern ML detection method. Maslan [4] use NB, KNN, SVM, Random Forest with live dataset to capture packet using Wireshark in application layer. Koroniotis [16] use Decision Tree, ANN and NB with NSL-KDD, UNSW-NB15 and BoT-IoT for finding standardization and common specification of detection. Lima[3] use Random Forest with UNSW and NetFlow dataset to create smart detection system to detect DoS and DDoS. De Donno [17] did a research on Mirai Variant dataset in order to detect the variation of DDoS in IoT Mirai and its future evolution in this field.

Table 1 - Summary of related work on ML-based detection techniques

Index	References	Feature Types	Dataset	Classification Algorithm	Accuracy (%)
1.	Akanji et al. [12]	DDoS	NetFlow	Genetic Algorithm,SVM	99.89%
2.	Awan et al. [13]	DDoS	SparkML	RF, KNN, MLP, Scikit	99.5%
3.	Swe et al. [14]	DDoS	NetFlow, CSE-CIC-IDS 2017, 2018	RF, KNN, MLP, PART	99%
4.	Wang et al. [15]	DDoS	NSL-KDD	Dynamic MLP, SBP-MLP	92%
5.	Maslan et al. [4]	DDoS	Live Dataset	NB, KNN, SVM, RF	98.70%
6.	Koroniotis [16]	DDoS	NSL-KDD, UNSW-NB15, UNSW BoT-IoT	DT, ANN, NB	99.45%
7.	Lima Filho et al. [3]	DDoS	CIC-DoS, CICIDS2017, CSE-CIC-IDS2018	RF	96%
9.	De Donno et al. [17]	DDoS	Mirai Variant	-	99.9%

2.6 Feature Selection Technique

The study applied two feature selection techniques which is Information Gain (IG) and Gain Ratio (GR). IG is built to predict variables by reducing entropy after cleaning and splitting. GR lessen the bias of IG using Intrinsic Information (II). After reducing entropy through cleaning and splitting, GR will predict the variables. This value shows how much a feature contributes to a change in the model's output.

3. Methodology

There are seven phases applied in this research which are data preparation, data filtering, data cleaning, feature selection, classification, performance evaluation and data presentation as shown in Figure 1 and Figure 2. The research framework in Figure 3.2 shows that two datasets will be used as the input of extraction. Only one dataset will be chosen as the main input and extraction of data according to the performance of the model evaluated by the metrics. Two datasets involved are NF Ton-IoT, and NF BoT IoT where the datasets will be train by different ML algorithm which are the Naïve Bayes, KNN, Decision Table and Random Forest. This project will be validate using parameter evaluation Accuracy, True Positive, True Negative, Precision and Recall. There are 5 phases of report writing and documentation of this research starting with problem identification, literature reviews, setting objectives and methodologies, choosing datasets and algorithms, running the dataset, processing and analyzing the data and report writing. The first steps analyze any related problem occurring during the analysis of the reading process. Findings in phase 1 which narrow down the scope of the problem to be improved is discussed in phase 2 in proper literature review format by comparing twelve research papers, journals and IEEE papers. Phase 3 describes the research objectives and methodology to be used for this research in detail. The chosen dataset and algorithm are discussed in phase 4 in order to analyses the behavior of the dataset in detail and recorded in phase 5. All the results will be analyzed and recorded for any future recommendation in technical writing format.

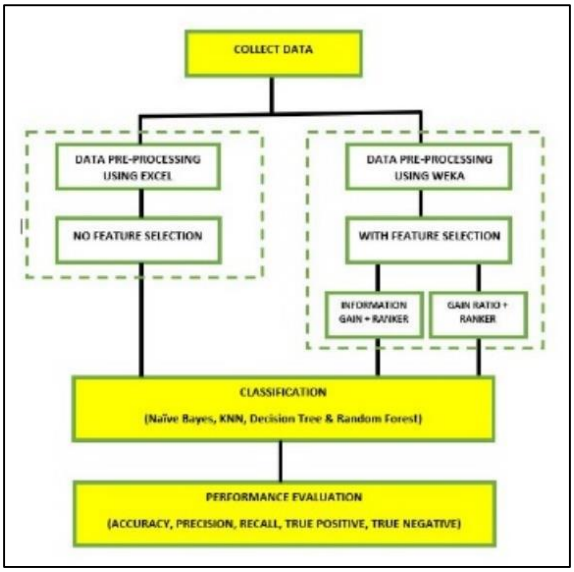

Fig. 1 - Research framework

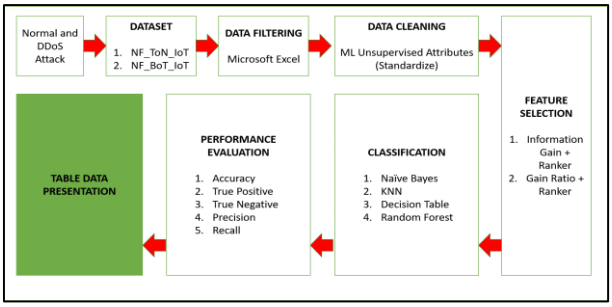

Fig 2 - Research methodology

3.1 Data Preparation

Two dataset is used in this study which are:

- Dataset 1: NetFlow ToN-IoT
- Dataset 2: NetFlow BoT-IoT

3.2 Data Preprocessing

The dataset contains missing value and redundant data is removed. Only 25000 of each normal and attack data selected from both datasets to be used in this experiment. Both datasets have 12 features and 2 class for detection.

3.3 Split into Training and Validation Data

The dataset is split into training and testing data to be used in 10-fold cross validation. The validation data is used to validate the performance with the unseen data to see if the model able to generalize well.

3.4 Feature Selection Algorithm

This research uses two feature selection techniques which are IG and GR to select the most important features value. The features are ranked from highest to lowest and the top 4, 8 and 12 are selected for training the ML model. The formulae to calculate IG is given in (1) and (2).

$$\text{Entropy} = -\sum P(x) \times log_2 P(x) \quad (1)$$

$$IG\ (X;\ Y) = H(X) - H\ (X \mid Y) \quad (2)$$

Gain Ratio attempts to lessen the bias of Information Gain on highly branched predictors by introducing a normalizing term called the Intrinsic Information (II). II is defined as the entropy of sub-dataset proportions. In other

words, it is hard to guess in which a randomly selected sample is put into. The more entropy being reduce after cleaning and splitting, the more we get the result. Through this filtering and sorting attributes, a Classification Tree is built to predict the variables (3).

$$II = -\left(\sum \frac{|D_j|}{|D|} * \log_2 \frac{|D_j|}{|D|}\right) \quad (3)$$

3.5 10-Fold Cross Validations

Dataset will be divided into 10 subsets where when 1 subset tested, the remaining 9 subsets will become training data. This step will be repeated until all the subset has been tested and trained with iteration 10 which resulting higher accuracy of testing than 70:30 split. 10-fold cross validation divides a data set into 10 subsets. Each time, one subset is used as the test set and the other nine as a training set. A second subset of data will be used as test data, and the remainder as training data. This is repeated 10 times. The average error is calculated across all 10 trials.

3.6 Validate The ML Model

To validate the performance of ML datasets, data that has been separated in the early phase is used. The classification algorithms used are Naïve Bayes, KNN, Decision Table and Random Forest.

The Naïve Bayes algorithm assumes that each feature makes an independent and equal contribution to the outcome. Equation 4 provides a way of calculating the probability of P(y|X) from P(X), P(Y), and P(X|y).

$$\frac{P(X|y)P(y)}{P(X)} \quad (4)$$

In KNN algorithm, a data point is categorized using the classification neighbors. KNN is a straightforward algorithm that sorts new information or instances based on similarities between them and all previously stored examples.

In Decision Table inputs are compared to rules, cases and test conditions. It is powerful in handling complex software testing and handling requirements. Evaluation can be identified quickly by True and False value.

In the RF algorithm, several decision trees are constructed, and predictions are derived from them. Trees are based on predefined attributes selected randomly. Classification is done by majority vote for each tree. RF construct number of decision tree on various sample and use them as majority vote.

3.7 Performance Evaluation

In order to test and validate the model performance based on the ML filtering and classification, parameter evaluation using confusion matrix Accuracy, True Positive, True Negative, Precision and Recall will be done. Figure 3 3 shows how the confusion matrix works to calculate the results.

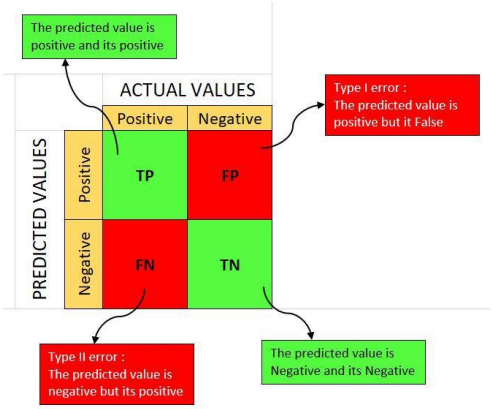

Fig 3 - Confusion matrix

Accuracy: The total number of correctly classified instances is the number of correct classifications of either normal or DDoS traffic in the dataset. The accuracy model is calculated using formulae (5).

$$\frac{TP+TN}{TP+TN+FN+FP} \quad (5)$$

Precision: Total number of normal traffic detected out of DDoS traffic. The precision model calculated using formulae in (6).

$$\frac{TP}{TP+FP} \quad (6)$$

Recall: Total number of normal traffic classified as Benign and DDoS traffic classified as Attack. The recall value of the model is calculated using formulae in (7).

$$\frac{TP}{TP+FN} \quad (7)$$

4. Result and Discussion

This section describes the findings of this study.

4.1 Implementation Tools

The experiment is done on a machine with AMD Ryzen 5 4600 H with Radeon Graphic processor with 12 CPUs of 3.00 GHz, 8GB RAM and 7GB NVIDIA Geforce GTX 1650 Ti. Weka Version 3.8.6 to run the experiment and manipulation of significant features.

4.2 Pre-Preprocessing

Dataset 1 is NF_ToN_IoT while Dataset 2 is NF_BoT_IoT. Each dataset was filtered using Microsoft Excel in order to reduce data to be tested and avoid software crashes. 50,000 test data with ratio 1:1 attack and benign was chosen, manipulated using ranked feature selection using two method selection Information Gain and Gain Ratio, and to be tested by 10-cross validation classifier using 4 different algorithm Naïve Bayes, KNN, Decision Table and Random Forest.

4.3 Ranking Using ML Methods Information Gain and Gain

Classification is one of the ways to manipulate the dataset to gain the best significant features for DDoS attack detection. To do this, both Dataset 1 and Dataset needs to have significant features to be compare and manipulate. Since each dataset has 12 features and 2 classes each other, this research will carry out three types of experiment which are applied ML algorithm towards the dataset without ranking, with Information Gain ranking and with Gain Ratio ranking attributes. Each experiment will undergo the 10-fold cross validation in order to get the train and test data. Each column and row will be folded and tested to test the accuracy using parameter evaluation.

Table 2 - Top 4,8,12 Features ranked by IG and GR

Index	Description	No Ranking	No Rank		IG		GR	
			D1	D2	D1	D2	D1	D2
1	IPv4 source address	IPV4_SRC_ADDR	1	1	1	7	4	6
2	IPv4 source port number	L4_SRC_PORT	2	2	8	4	1	4
3	IPv4 destination address	IPV4_DST_ADDR	3	3	9	6	3	5
4	IPv4 destination port number	L4_DST_PORT	4	4	3	8	8	10
5	IP protocol identifier byte	PROTOCOL	5	5	7	12	6	3
6	Layer 7 protocol (numeric)	L7_PROTO	6	6	2	10	5	8
7	Incoming number of bytes	IN_BYTES	7	7	10	11	7	11
8	Outgoing number of bytes	OUT_BYTES	8	8	9	3	9	7
9	Incoming number of packets	IN_PKTS	9	9	6	9	2	9
10	Outgoing number of packets	OUT_PKTS	10	10	12	2	10	12
11	Cumulative of all TCP flags	TCP_FLAGS	11	11	11	5	11	1

12	Flow duration in milliseconds	FLOW_DURATION_MILLISECONDS	12	12	5	1	12	2
13	Normal or Attack traffic	Label	13	13	13	13	13	13
14	Types of Attack	Attack	14	14	14	14	14	14

4.4 Result

Based on the Table 3,4 and 5, the best performance of the ML model for Dataset 1 without ranking is Naïve Bayes with accuracy of 97.51%. Best performance of Dataset 1 with Information Gain Ranking is Naïve Bayes using top 4 features with accuracy of 98.14%. Best performance of Dataset 1 with Gain Ratio Ranking is Naïve Bayes top 4 features with accuracy of 98.14%.

Table 3 - Performance result of ML models without ranking

Index	Classifier	D1 - NF ToN-IoT					D2 - NF BoT-IoT				
		Accuracy	TP Rate	FP Rate	Precision	Recall	Accuracy	TP Rate	FP Rate	Precision	Recall
1	NB	97.506	0.975	0.025	0.976	0.975	90.67	0.907	0.224	0.911	0.907
2	KNN	99.99	1	0	1	1	99.91	0.999	0.001	0.999	0.999
3	DT	100	1	0	1	1	99.88	0.999	0.002	0.999	0.999
4	RF	100	1	0	1	1	99.96	1	0	1	1

Table 4 - Performance result of ML models using dataset 1 and dataset 2 with IG

	Feature Selection IG + RANKER		NF ToN-IoT					NF BoT-IoT				
			Accuracy	TP Rate	FP Rate	Precision	Recall	Accuracy	TP Rate	FP Rate	Precision	Recall
1	IG + Ranker + Top 4	NB	98.136	0.981	0.019	0.982	0.981	86.388	0.864	0.349	0.882	0.864
	IG + Ranker + Top 8		94.47	0.945	0.055	0.95	0.945	88.766	0.888	0.286	0.9	0.888
	IG + Ranker + Top 12		97.506	0.975	0.025	0.976	0.975	90.672	0.907	0.224	0.911	0.907
2	IG + Ranker + Top 4	KNN	99.998	1	0	1	1	99.96	1	0	1	1
	IG + Ranker + Top 8		99.996	1	0	1	1	99.966	1	0	1	1
	IG + Ranker + Top 12		99.996	1	0	1	1	99.906	0.999	0.001	0.999	0.999
3	IG + Ranker + Top 4	DT	100	1	0	1	1	99.888	0.999	0.002	0.999	0.999
	IG + Ranker + Top 8		100	1	0	1	1	99.888	0.999	0.002	0.999	0.999
	IG + Ranker + Top 12		100	1	0	1	1	99.888	0.999	0.002	0.999	0.999
4	IG + Ranker + Top 4	RF	99.96	1	0	1	1	99.966	1	0	1	1

	IG + Ranker + Top 8		99.992	1	0	1	1	99.942	0.999	0.001	0.999	0.999
	IG + Ranker + Top 12		99.996	1	0	1	1	99.954	1	0	1	1

Table 5 - Performance result of ML models using dataset 1 and dataset 2 with GR

	Feature Selection GR + RANKER		NF ToN-IoT					NF BoT-IoT				
			Accuracy	TP Rate	FP Rate	Precision	Recall	Accuracy	TP Rate	FP Rate	Precision	Recall
1	GR + Ranker + Top 4	NB	98.136	0.981	0.019	0.982	0.981	86.672	0.867	0.335	0.881	0.867
	GR + Ranker + Top 8		97.302	0.973	0.027	0.974	0.973	93.836	0.938	0.153	0.942	0.938
	GR + Ranker + Top 12		97.506	0.975	0.025	0.976	0.975	90.672	0.907	0.224	0.911	0.907
2	GR + Ranker + Top 4	KNN	99.9988	1	0	1	1	98.906	0.989	0.027	0.989	0.989
	GR + Ranker + Top 8		99.994	1	0	1	1	99.966	1	0	1	1
	GR + Ranker + Top 12		99.996	1	0	1	1	99.906	0.999	0.001	0.999	0.999
3	GR + Ranker + Top 4	DT	100	1	0	1	1	98.872	0.989	0.027	0.989	0.989
	GR + Ranker + Top 8		100	1	0	1	1	99.888	0.999	0.002	0.999	0.999
	GR + Ranker + Top 12		100	1	0	1	1	99.888	0.999	0.002	0.999	0.999
4	GR + Ranker + Top 4	RF	99.96	1	0	1	1	98.912	0.989	0.027	0.989	0.989
	GR + Ranker + Top 8		99.998	1	0	1	1	99.966	1	0	1	1
	GR + Ranker + Top 12		99.996	1	0	1	1	99.958	1	0	1	1

The best performance of the ML model for Dataset 2 without ranking is Random Forest, with an accuracy of 99.96%. The best performance for Dataset 2 with Information Gain Ranking is Random Forest top 8 with an accuracy of 99.94%. The best performance for Dataset 2 with Gain Ratio Ranking is KNN top 12 with an accuracy of 99.94%.

5. Conclusion and Future Works

This research is done in order to find the best feature selection for DDoS bot attack detection. Both datasets, including 12 features, has gone through several feature selection techniques and algorithm in Machine Learning environment. The datasets with 12 features were selected to identify the best feature subset. This study performs feature selection using the ML Information Gain and Gain Ratio method to select and sort the features using Ranker to find the most important and

effective detection. Four classifiers have been used to evaluate the performance of the models, which are the Naive Bayes, K-Nearest Neighbour (KNN), Decision Table and Random Forest to show the performance of each dataset in their Accuracy. It could be important to investigate other techniques to improve the feature selection in the future. Some of the related study which has been identified are (1) using the latest dataset to evaluate and compare the performance and new features of the DDoS bot attack, (2) using better machine and hardware capacity to test and evaluate large dataset for training and testing, (3) use a combination of selection techniques and methods to choose the best 10 feature to improve computing and processing times for detection.

Acknowledgement

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) and Book Hack Enterprise through SEPADAN RE-SIP (vot M071).

References

[1] Khan, N. A., Brohi, S. N., & Zaman, N. (2020). Ten Deadly Cyber Security Threats Amid COVID-19 Pandemic. *TechRxiv Powered by IEEE*, *May*, 1-6. https://www.techrxiv.org/articles/Ten_Deadly_Cyber_Security_Threats_Amid_COVID-19_Pandemic/12278792

[2] Cristea, L. M. (2020). Current security threats in the national and international context. *Journal of Accounting and Management Information Systems*, *19*(2), 351-378. https://doi.org/10.24818/jamis.2020.02007

[3] Lima Filho, F. S. De, Silveira, F. A. F., De Medeiros Brito Junior, A., Vargas-Solar, G., & Silveira, L. F. (2019). Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning. *Security and Communication Networks*, *2019*. https://doi.org/10.1155/2019/1574749

[4] Maslan, A., Mohamad, K. M. Bin, & Mohd Foozy, F. B. (2020). Feature selection for DDoS detection using classification machine learning techniques. *IAES International Journal of Artificial Intelligence*, *9*(1), 137-145. https://doi.org/10.11591/ijai.v9.i1.pp137-145

[5] Alzahrani, S., & Hong, L. (2018). Generation of DDoS Attack Dataset for Effective IDS Development and Evaluation. *Journal of Information Security*, *09*(04), 225-241. https://doi.org/10.4236/jis.2018.94016

[6] Jithu, P., Shareena, J., Ramdas, A., & Haripriya, A. P. (2021). Intrusion Detection System for IOT Botnet Attacks Using Deep Learning. *SN Computer Science*, *2*(3). https://doi.org/10.1007/s42979-021-00516-9

[7] Sagar Dhanraj Pande, A. K. (2019). *A Review on Detection of DDOS Attack Using Machine Learning and Deep Learning Techniques*. *16*, 2035-2043.

[8] Odumuyiwa, V., & Alabi, R. (2021). DDOS Detection on Internet of Things Using Unsupervised Algorithms. *Journal of Cyber Security and Mobility*, *10*, 569-592. https://doi.org/10.13052/jcsm2245-1439.1034

[9] Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security. *IEEE Communications Surveys and Tutorials*, *22*(3), 1646-1685. https://doi.org/10.1109/COMST.2020.2988293

[10] Alothman, Z., Alkasassbeh, M., & Al-Haj Baddar, S. (2020). An efficient approach to detect IoT botnet attacks using machine learning. *Journal of High Speed Networks*, *26*(3), 241-254. https://doi.org/10.3233/JHS-200641

[11] Malinowski, E., Zimányi, E., Joseph, S. K., Warehouse, D., Inmon, B., Analytical, O., Olap, P., Gatziu, S., Vavouras, A., Nilsson, A. A., & Merkle, D. (2019). About the Tutorial Copyright & Disclaimer. *Data Vault 2.0*, *January 1999*, 1-15. https://doi.org/10.1007/978-3-322-94873-1

[12] Akanji, O. S., Abisoye, O. A., & Iliyasu, M. A. (2021). Mitigating Slow Hypertext Transfer Protocol Distributed Denial of Service Attacks in Software Defined Networks. *Journal of Information and Communication Technology*, *20*(3), 277-304. https://doi.org/10.32890/JICT2021.20.3.1

[13] Awan, M. J., Farooq, U., Babar, H. M. A., Yasin, A., Nobanee, H., Hussain, M., Hakeem, O., & Zain, A. M. (2021). Real-time ddos attack detection system using big data approach. *Sustainability (Switzerland)*, *13*(19), 1-19. https://doi.org/10.3390/su131910743

[14] Swe, Y. M., Aung, P. P., & Hlaing, A. S. (2021). A slow ddos attack detection mechanism using feature weighing and ranking. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 4500-4509.

[15] Wang, M., Lu, Y., & Qin, J. (2020). A dynamic MLP-based DDoS attack detection method using feature selection and feedback. *Computers and Security*, *88*. https://doi.org/10.1016/j.cose.2019.101645

[16] Koroniotis, N. (2020). *Designing an effective network forensic framework for the investigation of botnets in the Internet of Things*. *March*.

[17] De Donno, M., Dragoni, N., Giaretta, A., & Spognardi, A. (2018). *DDoS-Capable IoT Malwares: Comparative Analysis and Mirai Investigation*. https://doi.org/10.1155/2018/7178164