Machine Learning-Based Performance Analysis for IoT Attacks in IDS

Myint Soe Khaing
Cisco Lab, UCSY
University of Computer Studies, Yangon
Yangon, Myanmar
myintsoekhaing@ucsy.edu.mm

Zin May Aye
Cisco Lab, UCSY
University of Computer Studies, Yangon
Yangon, Myanmar
zinmayaye@ucsy.edu.mm

Thazin Tun
Cisco Lab, UCSY
University of Computer Studies, Yangon
Yangon, Myanmar
thazintun@ucsy.edu.mm

***Abstract*— Internet of Things(IoT) security is one of the main issues when executing and creating IoT platforms. The significant increment of the IoT devices in smart homes and other smart infrastructure make numerous attacks on these devices. With new and interesting attacks equipped for trading off the IoT platforms, intrusion detection and forensic systems really should be created. Security systems, for example, cryptography and validation are difficult to apply obliged IoT devices and organizations. With predominant innovations like the IoT, Cloud Computing, and Social Networking, a lot of organization traffic and information are created. Subsequently, there is a requirement for Intrusion Detection Systems that screen the organization and break down the incoming traffic powerfully. IDS assumes a significant function as a high-security answer for intrusion detection in IoT networks. Building an efficient network-based IDS, feature selection is an essential step as irrelevant and redundant features may adversely affect the classification performance of the system. The proposed system implements using Weka Tool based on BoT-IoT Benchmark dataset. The aims of the system to detect attack utilizing three different machine learning algorithms such as J48, Hoeffding Tree, and Naïve Bayes (NB). The solution for this problem may be provided by calculating precision, recall and, F1-score based on confusion matrix.**

Keywords— Botnets, Internet of Thing (IoT), J48, Hoeffding Tree, Naive Bayes (NB), Decision Tree.

I. INTRODUCTION

The IoT is a network of interconnected normal articles called "things" that have been extended with a little extent of figuring limits. As of late, the IoT has been impacted by a wide scope of botnet works out. As botnets have been the purpose behind real security possibilities and cash related damage all through the long haul, existing Network legal strategies can't perceive and follow current refined methods for botnets. Through different evaluations beginning late have talked about the utilization of Machine Learning (ML) plans in intrusion detection issues, little idea has been given to the identification of attacks unequivocally in IoT networks.

Machine Learning can be portrayed as a canny contraption's ability to change or mechanize an information-based state or conduct, which is seen as an essential bit of an IoT plan. ML can prompt obliging information from information made by gadgets or individuals, and ML algorithms are used in errands, for example, backslide, and order. Moreover, in an IoT community, ML may be used to provide protection agencies. The usage of ML in attacks detection issues is reworking into an intensely searched after concern, and ML is being used an increasing number of unique programs inside the community safety subject. Through various assessments inside the composing have used ML systems to find the nice methods to deal with recognizing assaults, simply limited research exists on gainful detection techniques realistic for IoT conditions.

An IDS or network Intrusion Detection System has been advanced that is prepared for perceiving an extensive scope of community eccentricities or assaults in the entrance conditions. The IDS is ready in the network that it guarantees, and it assembles community packets unpredictably likewise as a packet analyzer or network sniffer. Intrusion detection can be contributed by deploying data mining. Tree-based classifiers have a non-linear and hierarchical methodology that can be utilized for non-parametric and all-out data. They have brilliant adaptability in data investigation as they can uncover the hierarchical structure of the free factors.

In this examination, we expect to add to the structure by evaluating unmistakable machine learning algorithms that can be applied to rapidly and correctly recognize IoT network attacks. The main objective of the proposed system is to predict the malicious attacks with categories using BoT-IoT dataset and to know the workflow of IDS and analyze the system performance using three different Machine Learning algorithms. This paper is to compare the performance of classification in three different machine learning algorithms using the BoT-IoT dataset. The solution for this problem may be provided by calculating precision, recall and, F1-score based on confusion matrix.Classification is a technique that organizes data of a given class. For example, J48, Hoeffding Tree, and Naïve Bayes to analyze BoT-IoT dataset utilize Weka (Waikato Environment for Knowledge Analysis) Tools. The classification tab in Weka Explorer is utilized for classification where a few classification methods or algorithms like Bayes, trees, functions, rules, and Meta are available. The performance analysis of ML algorithms using Weka tool is presented in this paper.

II. RELATED WORK

IoT primarily based systems required an assured and short correspondence interface among the implanted contraption and the internet. The investigation of interference detection in IoT has gotten a super deal of concept because of the inadequacies and dangers available in IoT networks. These deficiencies required new techniques of interference detection to fill these openings. Therefore, researchers have been conducted a lot of research about this problem for several years. Sun et al. [2] proposed a circulate-primarily based IDS using TCP circulate as the rule of thumb model to understand and bunch malignant practices using Benford's regulation. Their research indicates that every attack has a splendid model, and the use of these models, they safely type the standard movement and anomalous circulation. Various intrusion detection strategies rely upon packet research and aid overwhelming in a quick community. Gupta et al. [3] proposed they have executed and given a similar examination

of various data mining techniques. The network attacks can be perceived suitably by applying direct backslide with 80% accuracy and 67.5% accuracy is developed by applying a semi-administered methodology for example k-deduces gathering approach. Goeschel [1] proposed an epic methodology is proposed in which SVM, decision trees, and Naïve Bayes are assembled to redesign IDS performance. For the most part, accuracy was over 99.62% with a false-positive rate of 1.57% anyway the last stage FPR was 4.29% for instance exceptionally higher than the underlying two phases.

Most people of the significant works utilize the NSL-KDD and KDD datasets, which are out of date and the outstandingly confined useful impetus for a front line IDS. Each type and harmful network site visitors have changed with the aid of and massive since 1999 when these datasets were made and the results got using them are of a confined worth usually. To beat a couple of inadequacies of as of late proposed strategies, like low recognition of an exceptional assault, misclassification of attacks, and time overhead, we propose assorted classifier models, specifically, J48, Hoeffding Tree, and Naïve Bayes. Besides, we use the BoT-IoT dataset to survey their presentation in distinguishing network intrusions and we contrast it and different ML techniques proposed by past masters.

III. Intrusion Detection System

Most IDSs have an average structure that joins: (1) an information-gathering module accumulates information, which maybe contains verification of an attack, (2) an investigation module recognizes an attack in the wake of setting up that information, and (3) a framework for uncovering an attack. In the information gathering module, the data information of each bit of IoT systems can be collected and reviewed to find a common lead of communication, appropriately recognizing harmful direct toward the starting stages. The analysis module can be realized using various methods and techniques, regardless, ML/DL based strategies are more sensible and winning for information evaluation to learn kindhearted and unusual direct reliant on how IoT gadgets and systems interface with one another in IoT conditions. Moreover, ML/DL algorithms can envision the new attacks, which is consistently not equivalent to past attacks, since ML/DL algorithms can astutely anticipate future dark attacks through picking up from existing bona fide models [5]. Figure 1 shows the parts of average IDS dependent on ML algorithms.

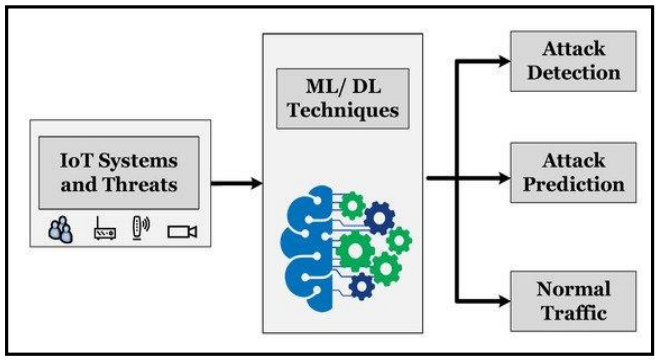


Fig. 1. Role of Deep learning and Machine Learning techniques based for IOT system.

Intrusion detection algorithms are arranged into two techniques: misuse detection and anomaly detection. Misuse detection algorithms recognize attacks dependent on acknowledged attack marks. They are valuable in perceiving known attacks with the least slip-ups. Regardless, they cannot see starting late made attacks that don't have comparable properties to the known attacks. Then again, quirk detection algorithms investigate conventional traffic and profile normal traffic plans. The anomaly detection technique depends upon the hypothesis that the attacker lead contrasts from that of a normal client. They organize traffic as an attack if the characteristics of the traffic are a long way from those of common traffic plans. IDSs can be huge for new attack plans. They are not as persuading as misuse detection models in the detection rate for known attacks and false-positive rates, which is a degree of misclassified normal traffic [4].

IV. Overall System Flow

In this paper, the overall system flow is illustrated in Fig. 2. It shows the overall system flow for this investigation beginning from the dataset and finishing with the evaluation of the different ML classifiers for attack analysis.

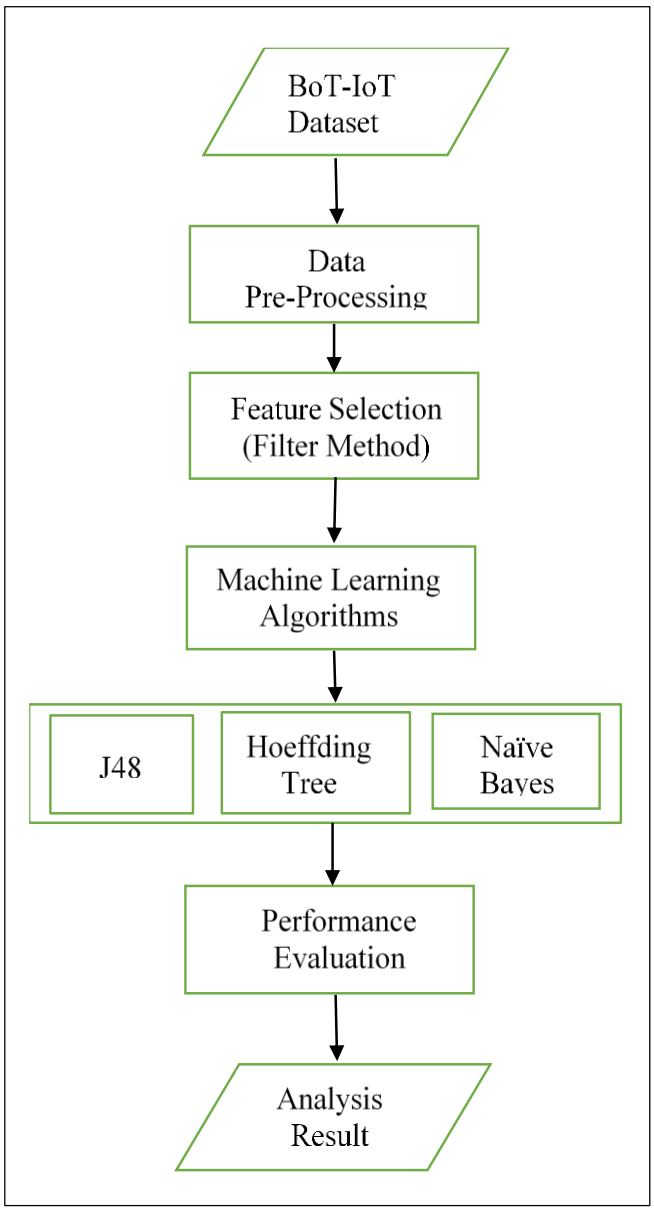


Fig. 2. The Overall System Flow

V. DATASET

The BoT-IoT dataset incorporates more than 72,000,000 data concocted on 74 documents, every line having 46 features. We make use of the edition proposed by way of Koroniotis et al., which is a spread of training and testing with 5% of the entire dataset. The BoT-IoT dataset [8] became made in the Cyber Range Lab of the Australian Center for Cybersecurity (ACCS). We utilized 177,569 instances (177,546 attack instances and 23 ordinary instances) from this dataset for performing the feature selection, which incorporates a sum of 46 features. Be that as it may, we picked the 43 features since one component is sequence ID and two features are another subclass highlight. This dataset has essentially three sorts of attacks that depend on botnet situations, for example, Probing, DoS, DDoS, and Information Theft [13].

VI. DATA PRE-PROCESSING

Data preprocessing is a data mining technique that includes changing raw data into a reasonable format. Real data is frequently deficient, conflicting, or potentially ailing in specific practices or drifts, and is probably going to contain numerous blunders. Data preprocessing is a demonstrated technique for settling such issues. Data preprocessing plan raw data for additional preparation. Data preprocessing is utilized in database-driven applications, for example, client relationships with the executives and rule-based applications (like neural organizations). Data pre-handling incorporates cleaning, Instance selection, normalization, transformation, feature extraction, and selection, etc.

A. *Data cleaning*

Data cleaning is a process of preparing the input data by removing incomplete, duplicated, irrelevant, improperly formatted data and fixing/smoothing the noisy data from the dataset. Various handling methods are used to play out every one of these errands, where every method is explicit to the client's inclination or issue set. Beneath, each task is clarified as far as the techniques used to conquer it.

B. *Data reduction*

Data reduction is a technique of reducing the volume of data capacity to reduce cost and increase storage data efficiency. There are many techniques to reduce the volume of data and these are described as follows:

- Missing values ratio:: Attributes that have more missing qualities than a limit are taken out.
- Low variance filter: It calculates variance from each numerical column and removes these columns which have variance value lower a given threshold value.
- High correlation filter: Standardized attributes which have a connection coefficient in excess of an edge are likewise eliminated since comparable patterns mean comparable data is conveyed. The relationship coefficient is typically determined utilizing statistical techniques, for example, Pearson's chi-square worth and so on.
- Principle Component Analysis (PCA): PCA is a dimensionality reduction technique which is used to reduce the large data sets' dimension by transforming a large variable set into smaller one which still has most information in large set.

C. *Data transformation*

Last advanced of data preprocessing phase is to transform the data into a proper form for applying machine learning model. The following techniques are empowered in data transforming process.

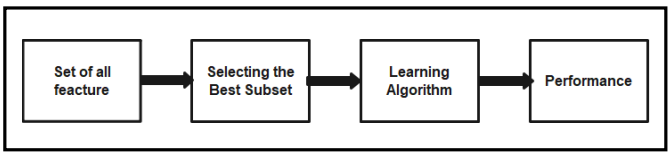
- Smoothing
- Feature Construction: New attributes/features are built from the given arrangement attributes.
- Aggregation: This task is applied to the given arrangement attributes to new attributes by concocting.
- Normalization: Each attribute value is scaled between more modest reach for example 0-1 or -1-1.
- Discretization: Raw estimation of numeric value attributes are supplanted by conceptual or discrete spans, which can consequently be additionally coordinated into more significant level stretches.
- Concept hierarchy generation for nominal data: Values for nominal data are summed up to higher-request concepts [17].

VII. FEATURE SELECTION

Feature selection (FS) or belongings choice is an example of selecting a subset of informational capabilities from the whole set. FS methods are applied to locate stimulated elements and crash insignificant, inconsequential qualities from data that doesn't affect the accuracy of farsighted models, on the off chance that they are melded or not, or may actually rot the accuracy of the model.

A. *Filter Method*

The filter method applies static measures to figure scores for every feature. A feature is both picked and discarded from the dataset situation to the score of every feature. For instance, information gain, Chi-squared test, and correlation coefficient



score.

Fig. 3. Filter Method

B. *Wrapper Method*

Wrapper methods take after a requesting issue, in which capabilities are set up in specific mixes, mentioned, and stood, separated from different mixes. A keen model is used which gives out a score reliant on model accuracy for evaluation of capabilities. Searching can be stochastic, for example, random hill-climbing algorithm, or heuristics, as ahead and in banter

passed to add and crash features. as an example, a recursive feature end algorithm.

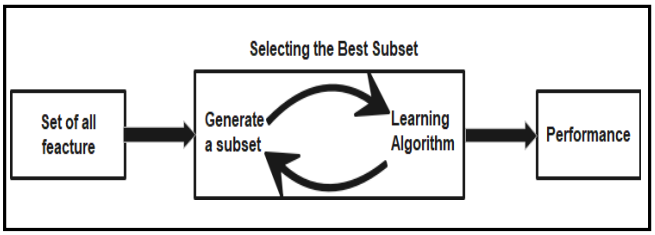


Fig. 4. Wrapper Method

C. *Embedded Method*

The embedded method checks every feature that develops the accuracy of the model at the same time as the model is being created [10].

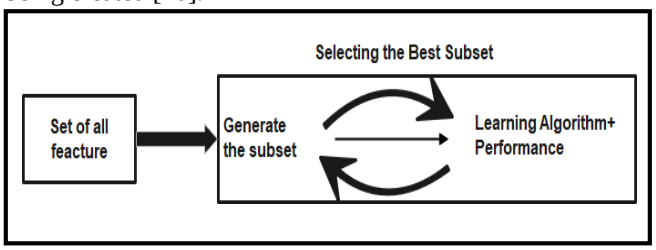


Fig. 5. Embedded Method

VIII. Machine Learning Methods

We utilized the BoT-IoT dataset to evaluate three simple ML classifiers: J48, Hoeffding Tree, and Naïve Bayes (NB). While picking these classifiers, the emphasis is on uniting well- known algorithms with various attributes. In this specific situation, the algorithms used are immediately reviewed in the accompanying segments.

A. *Decision Tree Algorithm*

Decision Tree algorithms are one of the used algorithms to settle classification wherein algorithms kind information into classes, like if an occasion is an attack. Decision trees contain nodes, branches, and leaves in which every center factor offers as a trickster contains and each branch gifts normally or decision, and each leaf gives ultimately. In applying the DT, it can be viewed as a motion of yes/no requests that are implemented with records to bring about a predicted class.

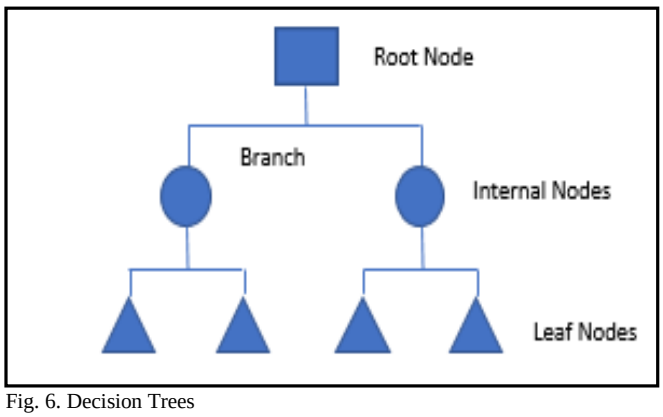


Fig. 6. Decision Trees

There are various algorithms achieved from the decision tree algorithm. Some of them are ID3, C4.5, J48, and many others the problem with the ID3 algorithm is that the data might be overfitted.C4.5 is an overhauled type of ID3 and it handles the overfitting difficulty. J48 is an open-wellspring of C4.5.

B. *J48 Classifier*

The J48 is utilized to paint distinct packages and perform specific consequences of the classification J48 algorithm is a champion among other AI algorithms to take a gander at the data categorically and continuously. The J48 decision tree can administer express qualities, lost or missing property evaluations of the data, and fluctuating brand name costs. Precision can be reached out by pruning (Venkatesan, 2015). The Algorithm is

- *Stage 1*: The leaf is named with an equivalent class if the models have a spot with a comparable class.
- *Stage 2*: For each quality, the ability data can be figured, and getting the facts will be taken from the test on the brand name.
- *Stage 3*: Finally, the quality excellent could be picked relying upon the modern-day decision parameter [7].

C. *Hoeffding Tree*

Hoeffding trees (HT) are first proposed by way of Hulten et al. (2001). One of the crucial figuring's for stream facts requests is the HT algorithm. It is a consistent, at whatever point decision tree (DT) enlistment calculation that is ready for learning from colossal information streams expecting that the vehicle-making models don't change over the long haul. It produces decision trees that take after the common get-together learning method. HT and DT are asymptotically related. A little model data would adequately be to pick an ideal splitting characteristic in a Hoeffding tree maintained numerically subject to the Hoeffding bound. The algorithm of Hoeffding Trees has guaranteed unrivaled which isn't essential with the other decision tree understudy. Decision trees are standard methodologies of data mining that are prepared for isolating hid data from datasets including nominal and numerical data.

$$\varepsilon = \sqrt{\frac{R^2(1-\delta)}{2N}} \quad (1)$$

HT algorithm depends upon an immediate thought that somewhat model can be regularly adequate to pick an ideal parting brand name. Mathematically, it is shown that the HT algorithm uses Hoeffding bound. To comprehend the centrality of Hoeffding bound commonly several speculations are made.

D. *Naïve Bayes*

The presence of one specific feature doesn't influence the other. Thus it is called naive. Naive Bayes is where the machine learning model is assembled dependent on the probability to be utilized in classification utilizing the Bayes theorem. For instance, by applying the Bayes theorem, we can figure the probability of an attack to happen when an occasion has happened.

Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

Utilizing Bayes theorem, we can discover the likelihood of an occurrence, given that B has happened. Here, B is the proof

and A is the hypothesis. The supposition made here is that the predictors/features are autonomous.

IX. Performance Analysis Of Machine Learning Algorithm

The confusion matrix is constructed on three diverse algorithms. Such as J48, Hoeffding, and Naive Bayes from Tables II-IV. A confusion matrix is a procedure for summing up the performance of the classification algorithms. Precision, Recall, and F-measure can be determined by utilizing the data given in a confusion matrix.

TABLE I. CONFUSION MATRIX

Actual	Predicted	
	Negative	***Positive***
Negative	∑TN	∑FP
Positive	∑FN	∑TP

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \quad (3)$$

$$Recall = \frac{\text{True Positive}}{\text{True Positive+False Negative}} \quad (4)$$

$$F\ measure = \frac{2(\text{True Postitive})}{2\text{True Positive+False Positive+False Negative}} \quad (5)$$

The confusion matrices of the three machine learning algorithms are presented in Tables II-IV. The Weka tool is applied to the BoT-IoT dataset with 10-fold cross-validation. According to the tables, the J48 algorithm offers satisfying accuracy is compared with other ML algorithms in classifying IoT attacks.

TABLE II. J48 CONFUSION MATRIX FOR CATEGORY

Normal	Reconnaissance	DDoS	DoS	Theft	Prediction/ Actual
20	3	0	0	0	**Normal**
0	17422	0	0	0	**Reconnaissance**
0	0	40026	0	0	**DDoS**
0	0	0	119997	0	**DoS**
0	0	0	0	78	**Theft**

TABLE III. HOEFFDING TREE CONFUSION MATRIX FOR CATEGORY

Normal	Reconnaissance	DDoS	DoS	Theft	Prediction/ Actual
12	0	0	11	0	**Normal**
1	17253	0	168	0	**Reconnaissance**
0	0	40008	18	0	**DDoS**
1	0	0	119996	0	**DoS**
0	1	0	0	77	**Theft**

TABLE IV. NAÏVE BAYES CONFUSION MATRIX FOR CATEGORY

Normal	Reconnaissance	DDoS	DoS	Theft	Prediction/ Actual
18	5	0	0	0	**Normal**
125	17146	0	151	0	**Reconnaissance**
18	0	40004	4	0	**DDoS**
1	2	7	119985	2	**DoS**
0	0	0	0	78	**Theft**

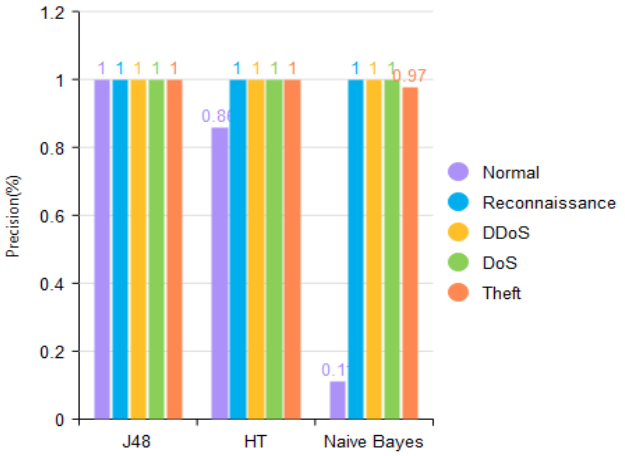


Fig. 7. A comparison of precision of J48, Hoeffding Tree (HT), and Naïve Bayes

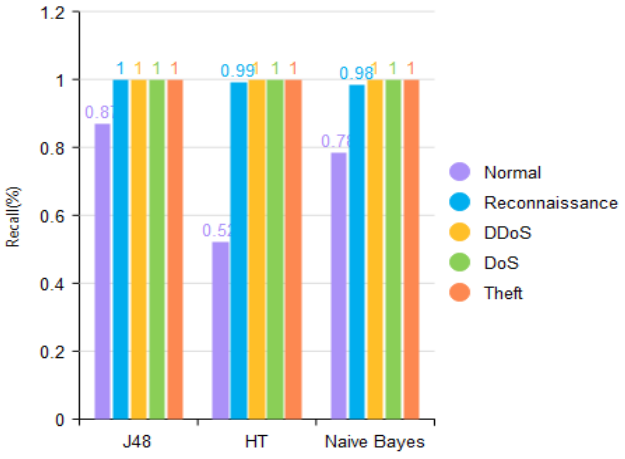


Fig. 8. A comparison of recall of J48, Hoeffding Tree (HT), and Naïve Bayes

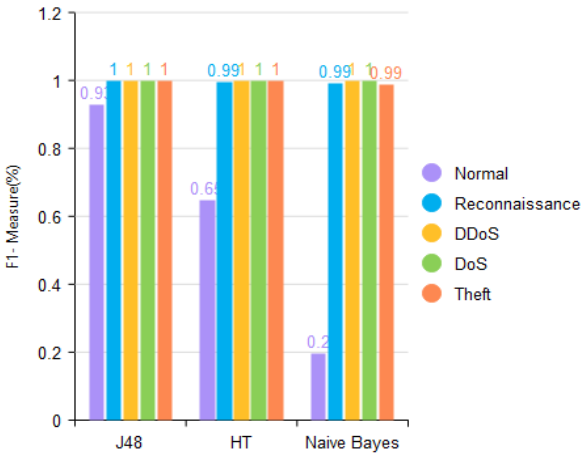


Fig. 9. A comparison of F1-measure of J48, Hoeffding Tree (HT), and Naïve Bayes

X. Conclusion

The performance analysis of the system is performed using Weka tool based on machine learning methods. The botnet attacks are the most recent attack on the IoT environment. It is needed to protect the IoT devices from these kinds of attacks. However, there are difficult to implement the attack detection system on IoT devices because they have very limited resources. Their performance is measured by ascertaining different evaluation measurements in particular precision, recall, f-measure. After that, we saw that a large portion of the algorithms utilized in this paper has great accuracy regarding classifying between the attack packets and normal packets and could be extremely useful in recognizing an interruption. J48 offers the best accuracy among other algorithms. J48 could be exceptionally helpful in the location of noxious exercises in the organization in an efficient way when we use classification algorithms for finding attacks. Therefore, this model could be utilized as an Intrusion Detection System for IoT development. The proposed system implements using Weka Tool based on Bot-IoT Benchmark dataset. The aims of the system to detect attack utilizing three different machine learning algorithms such as J48, Hoeffding Tree, and Naïve Bayes (NB).The results show that Decision Tree J48 Classifier is the best for classifying Botnet and normal network traffic. For future work, we will be applying other classification procedures for IoT security to meet the desired analyses related to the applied data set and proceeds to find out the proper model for IoT based IDS.

References

[1] D. Gupta, S. Singhal, S. Malik, and A. Singh, “Network intrusion detection system using various data mining techniques,” International Conference on Research Advances in Integrated Navigation Systems (RAINS): IEEE, pp. 1–6, 2016.

[2] I Ullah, and QH. Mahmoud. "A filter-based feature selection model for anomaly-based intrusion detection systems." In 2017 IEEE International Conference on Big Data (Big Data), pp. 2151-2159. IEEE, 2017.

[3] K. Goeschel, “Reducing false positives in intrusion detection systems using data-mining techniques utilizing support vector machines, decision trees, and naive Bayes for off-line analysis,” SoutheastCon 2016: IEEE, pp. 1–6, 2016.

[4] M. Arjunwadkar Narayan., and TJ. Parvat. "An Intrusion Detection System,(IDS) with Machine Learning (ML) Model Combining Hybrid Classifiers." connections 1 (2015):

[5] M. A Al-Garadi, A. Mohamed,Al-Ali, A.; Du, X.;Guizani, M. A survey of machine and deep learning methods for internet of things (IoT) security. arXiv 2018, arXiv:1807.11023.

[6] M A Ferrag, L Maglaras, A Ahmim, M Derdour, and Helge Janicke. "RDTIDS: Rules and Decision Tree-Based Intrusion Detection System for Internet-of-Things Networks." Future Internet 12, no. 3 (2020): 44.

[7] N.SaravanaN, Dr.V.Gayathri "Performance and Classification Evaluation of J48 Algorithm and Kendall’s Based J48 Algorithm (KNJ48)". International Journal of Computer Trends and Technology (IJCTT) V59(2):73-80, May 2018. ISSN:2231-2803.

[8] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset,” 2018.

[9] N. Koroniotis, N Moustafa, E Sitnikova, and J Slay. "Towards developing network forensic mechanism for botnet activities in the iot based on machine learning techniques." In International Conference on Mobile Networks and Management, pp. 30-44. Springer, Cham, 2017.

[10] N. Sultana, N Chilamkurti, W Peng, and R Alhadad. "Survey on SDN based network intrusion detection system using machine learning approaches." Peer-to-Peer Networking and Applications 12, no. 2 (2019): 493-501.

[11] P. K. Srimani., and Malini M. Patil. "Performance analysis of Hoeffding trees in data streams by using massive online analysis framework." International Journal of Data Mining, Modelling and Management 7, no. 4 (2015): 293-313.

[12] R. Patgiri, Udit Varshney, Tanya Akutota, and Rakesh Kunde. "An investigation on intrusion detection system using machine learning." In 2018 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1684-1691. IEEE, 2018.

[13] S. Yan Naung, Yaokai Feng, Paulus Insap Santosa, Rudy Hartanto, and Kouichi Sakurai. "Rule Generation for Signature Based Detection Systems of Cyber Attacks in IoT Environments." Bulletin of Networking, Computing, Systems, and Software 8, no. 2 (2019): 93-97.

[14] S. Yan Naung, Yaokai Feng, Paulus Insap Santosa, Rudy Hartanto, and Kouichi Sakurai. "Towards a Lightweight Detection System for Cyber Attacks in the IoT Environment Using Corresponding Features." Electronics 9, no. 1 (2020): 144.

[15] https://moredvikas.wordpress.com/2018/10/09/machinelearning-introduction-to-feature-selection-variable/selection-or-attribute-selection-or-dimensionality-reduction/

[16] https://www.google.com/search?client=firefox-b-d&q=j48+ algorithm +advantages

[17] https:// heartbeat. fritz.ai/ data- preprocessing- and- visualization-

[18] implications-for-your-machine-learning-model-8dfbaaa51423