International Journal of Research in Engineering, Science and Management
Volume 4, Issue 1, January 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

The significance of Machine Learning in Enhancing the Security for IoT networks to Detect Attack Signatures

Ashraf Siddiqui*
Aligarh University

***Abstract*: The Internet of Things (IoT) has contributed to many risks to defence and societal issues. Notwithstanding the social gains, IoT may jeopardise the protection and privacy of individuals and businesses at different levels. The most popular forms of attacks faced by IoT networks involve denial of service (DoS) and distributed dos (DDoS). Companies can use an accurate classification and identification model, which is not a simple job, to fight such assaults. This paper provides a model for the classification of many algorithms for machine learning, i.e. Random Forest (RF), k-Nearest Neighbors (KNN), and Naïve Bayes. The algorithms of the machine learning are used to identify assaults on the UNSW-NB15 dataset. The UNSW-NB15 involves regular network traffic and malicious traffic. The experimental findings indicate that RF and KNN classificators are the highest performers with a 100% accuracy (without injection), 99% (with 10% noise filter), and the Naïve Bayes classification provides the lowest outputs with 95.35% accuracy and 82.77% noise and ten% noise. Additional evaluation matrices, such as accuracy and reminder, also illustrate the utility of RF and KNN classification over Naïve Bayes.**

***Keywords*: Internet of Things, Security, Classification model, Machine Learning, Random Forest, k-Nearest Neighbors, Naïve Bayes.**

1. Introduction

Internet of Things (IoT) is a network of devices that allows these devices to share information directed towards different purposes [1]. Such devices include desktops, laptops, smartphones, and tablets. The inception of smart devices to the society was first done in 1982, where the first device to ever be connected to the Internet was a Coca-Cola Company vending machine. This machine kept stock of its commodities and kept inventory for the inputs and outputs. The machine also monitored the temperature of the drinks within the machine. The term IoT was coined by Kevin Ashton of Proctor and Gamble in 1999. However, the actual existence of aspects that verified this term came into existence in 2008. Figure 1 shows the concept of IoT [2].

IoT technologies are deployed in a number of areas, including customers, industry, industrial and infrastructure. For the commercial market, the IoT is incorporated into homes with such things as the presence of clever homes, i.e. homes which are capable of carrying out the most basic tasks that human interference originally demanded. Temperature monitoring and protection mechanisms and actions such as fire suppression using smoke alarms provide these features. IoT has now been implemented into healthcare programmes by the usage of mechanised collection and processing of medical records. As a product of the interaction of technology and the system, computers that can analyse the signs of individuals and promote illness detection are now used in the health field. The application of IoT is integrated in the production processes in the automotive field. The usage of human labour has been minimised and production productivity increased when robots pick over food manufacturing, packing and sealing. The usage of machines in manufacturing processes has contributed to higher productivity standards as opposed to the manual handling of other factory processes.

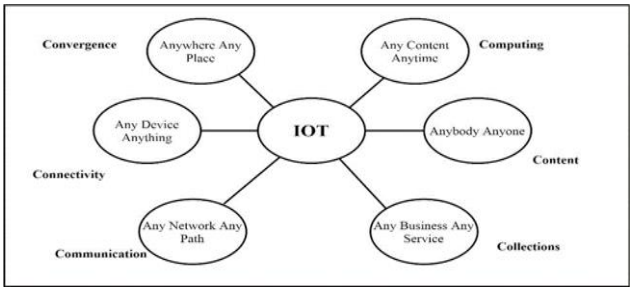

Fig. 1. The concept of IoT

In cultivation, the usage of IoT is often included within the industrial bracket. The usage of IoTs in automatic irrigation systems and the presence of environmentally managed greenhouses has been automated in agriculture, making it possible for almost anywhere in the world to grow [3]. The application of IoT in utilities has been used in aspects such as electricity storage. Energy use can be controlled by IoT. IoT is also used for environmental protection and surveillance of greenhouses [4]. However, the continuing engagement of IoT has contributed to the predisposition to multiple social problems. Regardless of the gains it has provided to society, IoT has jeopardised protection and privacy at multiple stages. Figure 2 demonstrates the problems of IoT security [5].

*Corresponding author: researchprojects48@gmail.com

Machine learning (ML) provides the programs with the ability to improve their performance with experience [6]. ML algorithms can be classified as supervised, unsupervised, and reinforcement learning. These categories can be used in areas weather forecasting, cluster identification and learning from mistakes, respectively. ML can assist the IoT arena by utilizing information on different security issues experienced and using it to make permanent solutions to the security threats for future preparedness. Also ML can help in detecting rare events or observations, i.e., the anomalies. Anomalies can raise suspicion because they are statistically different from other normal observations.

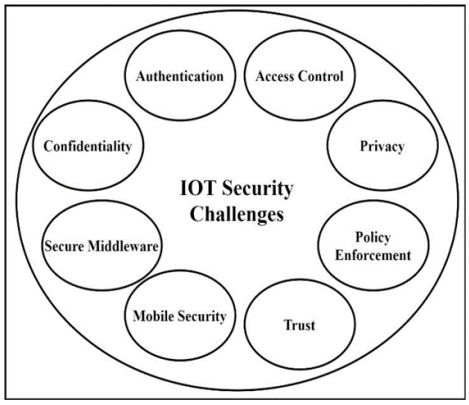

Fig. 2. IoT security challenges

The contribution of this work is to detect irregularities with the model of machine learning. A classification model has been developed to analyse the utility of a series of ML classification algorithms, namely Random Forest (RF), k-Nearest Neighbors (KNN) and Naïve Bays, based on a benchmark IoT dataset known as UNSW-NB15, and to determine the acceptable range of such algorithms in IoT environments for anomalies detection. A type of polling will also be used to enhance the estimation procedure.

The remainder of this article is structured accordingly. The history and literature review are discussed in Section 2. Section 3 offers information on the proposed classification model, in which numerous ML classifiers are used for model building. The findings and discussions of the concept implementation are discussed in Section 4. Section 5 ends the paper and proposes some options for potential work.

2. Background and Literature Review

IoT is a modern network that tracks the joint performance of machines within a single network. It requires the usage of a network link to operate devices without the requirement for human feedback or programming interference. Via IoT, smart vehicles are produced in community, homes are automated, manufacturing processes are operated, and other activities that no longer need manual input. However, IoT poses many obstacles among the advantages it has given to society. The usage of IoT is susceptible to numerous security issues which encourage malicious corruption of their systems by individuals or organisations. Since IoT networks are susceptible to attack because of such systemic loopholes that render it impossible to contain these insecurities.

A. The Classification of IoT

In order to combine multiple sensors and objects which can interact without the need of human interference, Alaba et al. [3] suggested an IoT solution. IoT is defined as including sensory devices which track and collect all kinds of information on the system as well as human social life. According to Hameed et al. [7], IoT is triggered by the interconnection between computers and networks owing to the technical development since the last century. The expectation of enhanced engagement would contribute to very large data collection.

In IoT, the device layer, the vision layer and the network layer are split into three levels. The framework layer in IoT is the highest layer that can be used by IoT end users. The layer of awareness is the layer to gather knowledge. It involves the nodes of perception and the network of perception. Finally, the network layer offers network transmission and a robust awareness system access setting. Digital knowledge has been, according to Pishva [8], a social infrastructure in our culture. IoT is viewed by the interconnection of computers with a similar network, the Internet. Online access is the main draw for IoT. As early as ten years ago, Japanese technologists led this with the development of Internet-enabled audiovisual devices [8]. IoT convergence has contributed to the development of the network economy. This is largely because the Internet has broken physical barriers which only restrict trade to a regional basis.

B. Security Issues and their Predisposing Factors in IoT

The sophistication of IoT has allowed the protection problems initially linked to the Internet to expand. Node entry, which is the IoT's fundamental functionality, is a predisposing factor in the problems facing this agency. A big contributing factor to IoT's safety issues is the absence of security to secure IoT devices from viruses and malware assaults. Only conventional networks provide cost limitation solutions to reduce the proliferation of threats and strengthen privacy security. In addition to the discussion of the traditional wireless network, Alaba et al. stressed the protection concerns of IoT. The traditional wireless network is the traditional Internet type which predisposes IoT entry. One essential aspect is the usage of Low Power and Lossy Networks (LLNs) that exposes the IoT to data loss due to the impersonation of nodes. Protection capabilities range from IoT and traditional networks. Since sensor nodes have low calculating power and low storage space, which is an important consideration in terms of IoT data traffic. IoT is often faced with protection problems such as fake and human assaults in the middle [3]. Both problems can collect network information and submit false data to network nodes.

IoT requires unified principles that can act as the foundation for protection challenges mitigation. Possible vulnerabilities cover devices and risks to the network. The future directions for IoT include the heterogeneity address, a form of IoT predisposing the related security risks [3]. Privacy is one of IoT's protection challenges. The combined identification of a person in IoT contributes to user profiling and monitoring.

Malicious actors may monitor IoT users and profile their experiences with their environment from the amount of knowledge that IoT produces. In order to protect users, the protection of IoT should provide a safe method of data transmission that reduces the chances of a person invading privacy, irrespective of the purpose. The lightweight cryptographic frame is another predisposing factor for IoT-affiliated vulnerability. Without a protection concession, IoT can be set to use less energy. At present, IoT uses less resources but is at the cost of stability, which puts the consumer at risk [7].

Frustaci et al. [9] also grouped IoT risks dependent on the IoT layer of understanding, transport and compliance. Depending on their understanding, including functional features of IoT such as sensors and nodes conducting data collection and perception, risks include physical invasion, impersonation, denial of service, routing attacks and attacks on data transit. Physical attacks require physical hardware disruption such as node modification or malware intrusion directly into the device. It consists in false identity through the usage of malicious nodes for impersonation. DoS assault utilises scarce network tools to circumvent legal users getting links to the network. On the other side, the routing attacks depend on manipulating the data routes during data processing and delivery. A data transit assault includes attacks like man-in-the-middle that intercepts and manipulates the data based on the hacker's well.

The transportation layer transmits the gathered information for the network. The security issues associated with the Transportation layer include routing attacks, DoS, and data transit attacks [9]. For the Application layer, the security issues include DoS, data leakage, and malicious code injection. Data leakage is the stealing of data based on its vulnerability. Frustaci et al. [9] highlighted the properties of trust in IoT and its importance which includes the certainty in collaboration, excellence in flexibility as well as the efficiency of the IoT. Trust in IoT has contributed to a loss of usage of many financial institutions and citizens who assume that their data is too fragile. By growing the confidence, IoT can be implemented and used internationally. Although IoT is closed and consumers are unable to apply protection software to smartphones, conventional networks will add antiviruses and other safety controls.

IoT will also only use lightweight algorithms that trigger high attachment affinities, as their goal is to combine higher protection with low system power. Traditional IT was managed by the consumer, while IoT gathers private user details automatically. Traditional IT devices are housed in closed areas, whereas IoT devices are situated in open environments.

Xiao et al. [6] stated that IoT has made it possible to connect the real universe with connectivity networks and their usage of the community in which we reside. The study outlined the need for fixing IoT-associated protection problems such as spoofing, interference, DDoS, jamming, scamming and malware.

C. Current Solutions to the Security Issues Facing IoT

In order to strengthen the security situation of IoT, its design should be improved to fulfil this reason. One approach is to create safe routes for data and knowledge exchange around the world. The separation of the malicious nodes used by hackers for their malicious behaviour is another prevention step. The framework should be revamped to allow malicious nodes to be identified and isolated. The device could also be strengthened to self-stabilize after an attack [7]. The safety protocol can direct the network to recovery without human interference. The defence of location privacy should be included in the framework to improve stability. The network should be configured to withstand the intrusion and ransomware attacks by identifying attackers as early as possible with respect to robustness and durability.

It should also facilitate swift recovery from potential errors arising from the attack. The device can be self-reliant and does not need human interference to retrieve or defend users from threats. Hameed et al. [7] examined DoS attacks that prevent the involvement of people from their network facilities as part of a cyber-attack to extract details from the culprit. This includes an effective resource counter measure and resource efficient monitoring of insider attacks. These properties may be used to resist DoS and DDoS. The issues currently confronting IoT are addressed in [7].

Pishva [8] demonstrated the Internet access limitations. Many of the threats associated with the Internet include database theft, privacy breaches and computer corruption. Any of these threats are created by Internet service providers, but are intended to improve service provision. These generated lapses, though, are used to extend company at the cost of customer victimisation and the use of attackers to execute their malicious behaviour. Another problem is the misuse of e-commerce secrecy. Buyers are already engaging with internet vendors of products and services. However, it has become an issue for Internet consumers when advertisers use the data of individuals to flood them with spam mail on the basis of their orders. One aspect that allows assaults on IoT users simpler is the absence of encryption capability. This is because the higher percentage of IoT-connected devices are our primary usable devices that are affordable and do not dictate the way costly tech is used. The presence of unknown persons in technology is also a factor in the predisposition of vulnerability in IoT. The system can only be linked to the Internet for typical users as used and turned off if no interactions are created, whereas intelligent devices are still connected to enable attacks.

One proposed security measure is the use of the United Home Gateway (UHG). This single pathway connects all household devices equipped with appropriate security measures. It prevents the access of the devices, as it is much easier to compromise them when they are directly connected to the Internet. Other proposed solutions include changing defaults passwords, disconnection of universal Plug and Play (PnP) features as they created security loopholes for an IoT device. The last proposition includes keeping the software up to date as this usually fixes security loopholes and bugs. By keeping a device's software updated, an individual stands a higher chance of protecting their devices from attack.

Table 1
Reviews of papers that use ML algorithms in the IoT security

Ref.	Attack types	Security techniques	ML techniques	Summary	Dataset used	Noise
[11]	Code confiscation	Dynamic monitoring	Naive Bayes	Discussing the concepts of code confiscation as a strategy that can help to create a secure mechanism for guaranteeing a secure architecture.	Constructed data	No
[12]	Cyber attacks	Efficient behavior approach	Averages One-Dependence Estimator (AODE)	Advancements in technology brought more challenges to the IoT field. Efficient capture behavior acts as a solution to the difficulties witnessed.	Constructed data	No
[13]	Cyber attacks	Blockchain approach	Reinforcement Learning (RL) algorithm	Dynamic access control policy is described in details and its abilities to provide the ultimate solution to security issues. The work uses the blockchain strategy and machine learning to create a solution.	Constructed dataset	No

D. Machine Learning in Developing IoT Security

To improve IoT protection, the best security activities for IoT security are methods such as authentication, access management, malware detection and stable loading. Authentication allows IoT devices to differentiate the root nodes and address attacks dependent on identification such as Sybil and spoof [6]. This avoids contact with hostile nodes that may avoid detecting assaults. The usage of an access management system is another solution for the protection problems in the IoT. Pass management prohibits unauthorised users from accessing IoT services. IoT should also only be utilised by limited legal entities who are given access to services on a particular computer. In addition, IoT devices may use the storage and measurement of IoT servers by including safe offloading techniques.

As stated in Section 1, ML provides machines with the opportunity to learn from previous experience and to enhance their output without a person's required feedback. The ML will support the IoT arena by the usage of knowledge on the numerous protection challenges and the lasting solution to the security risks to potential preparedness. This segment discusses recent papers utilising ML algorithms to strengthen the protection features of an IoT network. Andročec and Vrček carried out many experiments on the forms of IoT algorithms [10]. The authors noted that the protection of networks has been threatened by the high number of IoTdevices. They also reported that various malware cases target and destroy computers and systems on the Internet. The ML algorithms were used to support IoT security [10]. This portion includes an extensive analysis of the work of Andročec and Vrček [10] by the authors of the current article. Thirty-four recent experiments have been reported and reviewed to explain the significance of utilising ML to resolve IoT protection concerns. The results showed that further IoT protection ML tests have matured.

Table 1 summarises reviews of papers utilising ML algorithms for IoT protection, as well as other techniques. The attack styles, the protection techniques used, the ML techniques used by each analysis, a quick description, the kind of data set used, and whether or not the data set is loud as shown in Table 1. The examined documentation explains the significance of utilising ML algorithms to enhance IoT protection and thereby boost the attack detection mechanism. To our best understanding, however, no paper studies suggest that the usage of a voting algorithm to detect the attacks is significant. The voting algorithm blends various ML algorithms such that the prediction outcomes are enhanced and the security of IoT improved.

Section 3 presents the contribution of this work by introducing the proposed classification model. The proposed model detects the anomalies and hence increases the protection of the IoT environment. This can be achieved by examining the effectiveness of a set of ML classification algorithms, namely, Random Forest (RF), k-Nearest Neighbors (KNN) algorithm, and Naïve Bayes, on the IoT dataset, known as UNSW-NB15, to estimate the appropriate selection of such algorithms for detecting anomalies. The proposed model also applies a voting algorithm to produce an efficient predictive model and hence improving the estimation process over other ML classifiers. The details of the proposed classification model are given in Section 3 while its implementation results and discussion are presented in Section 4.

3. The Proposed Classification Model

This portion includes descriptions of the classification model suggested. Different ML classificators are added to create a model that can distinguish legally and illegally created traffic in various settings. The Weka platform shown in Figure 3 contains a set of ML data mining classifiers. Weka was used to incorporate the proposed model. Weka provides numerous methods for data regression, sorting, clustering, visualisation and the mining of association laws. The following paragraphs present the measures used, the ML classifiers used and the dataset benchmark used in the model suggested.

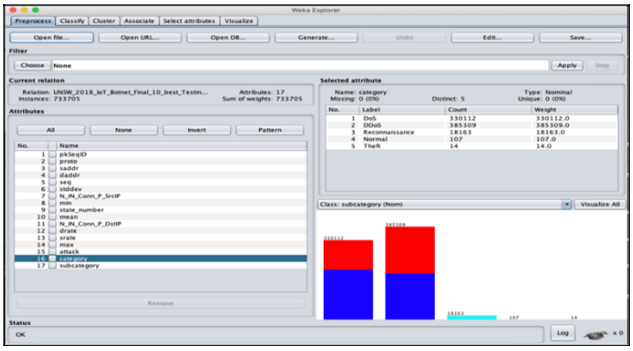

Fig. 3. The Weka platform

A. Evaluation Metrics

Confusion metrics will be used to evaluate the performance of the classifiers. Confusion metrics are commonly used in classification problems that have two or more types of classes.

The used confusion metrics in this research are accuracy, precision, and recall.

1) Accuracy

In classification problems, the accuracy is the number of correct predictions over all predictions made. It is a good measure when the target variable classes are nearly balanced.

2) Precision

Precision shows, out of all the positive observations, how many positive observations are predicted correctly. The higher number of correct predictions means the higher the performance of the classifier. Prediction can be given using the following equation:

$$\text{Precision}=TP/(TP+FP) \quad (1)$$

where TP represents true positive predictions and FP represents false positive predictions.

3) Recall

Recall shows, out of all observations in the actual class, how many positive observations are predicted correctly. As precision, the higher number of correct predictions is the higher the performance of the classifier. Recall can be given using the following equation:

$$\text{Precision}=TP/(TP+FN) \quad (2)$$

Where FN represents the false negative prediction.

B. Classification Algorithms

In the proposed model, the classification algorithms (i.e., classifiers) that will be used are Random Forest (RF), K-Nearest Neighbors (KNN), and Naïve Bayes. In addition to these algorithms, a voting algorithm will be applied in the proposed model. A voting method is an ML classifier that combines different models in order to produce an optimal predictive model that leads to improve the prediction results. The following subsections briefly present the used classifiers.

1) Random Forest

Random Forest (RF) that is commonly used because of its simplicity and the fact that it can be used for both regression and classification tasks. RF classifier creates several, but random, decision trees and merges them to produce a more accurate and stable prediction model [46].

2) K-Nearest Neighbours (KNN)

KNN algorithm supports both classification and regression. It stores a training dataset and conducts queries on the data set to locate the k most similar patterns to make predictions. KNN algorithm takes the category of the most similar items in the dataset and assign this category to the unlabeled instances.

3) Naïve Bayes

Naïve Bayes algorithm is a classification learning and statistical system. It is built using the training data set and estimates the possibility of each class in view of new instance characteristics.

The proposed classification model is being added to the UNSW-NB15 data collection, a modern IoT dataset covering a wide number of regular network traffic and malicious traffic instances. UNSW-NB15 is a dataset that acts as standard for the identification of malicious activities of the Network [49]. UNSW-NB15 was developed by the University of New South Wales to test new intrusion detection systems (IDS). A limit of 100 Go of raw network traffic has been obtained for the creation of UNSW-NB15 datasets. The dataset consists of 45 characteristics and consists of ten traffic types, one standard and nine distinct modes of attack. The workflow of the proposed classification model as seen in figure 4.

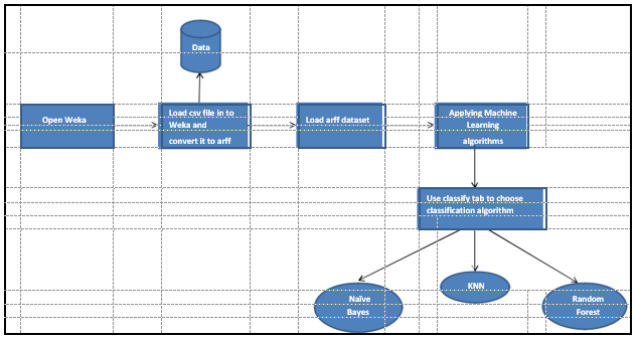

Fig. 4. The workflow of the proposed classification model

4. Results and Discussion

This portion describes the findings of the classification model suggested. A variety of ML classifiers, including KNN, Random Tree, Naïve Bayes and a voting system, were used to get results on the Weka platform. The suggested model is applied with a broad variety of regular network traffic and malicious traffic instances from a modern IoT dataset defined as UNSW-NB15. Different tests were performed using a collection of classifiers as seen in figures 5-12.

Total Number of Instances: 220111
Correctly Classified Instances: 220111 (100 %)
Incorrectly Classified Instances: 0 (0 %)
Mean absolute error: 0
Relative absolute error: 0.0078 %
Root mean squared error: 0.0016
Root relative squared error: 0.4813 %
Kappa statistic: 1

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Fig. 5. Random Forest classifier results summary

The number of correctly categorised instances was 100% for the RF and KNN classifiers based on experimental observations, whereas for the Naïve Bayes classifier, the percentage was about 95%. In comparison, the experimental findings demonstrate that the precision, accuracy, and recall parameters for the RF and KNN graders have the maximum values. Yet the same measurement criteria have the lowest Naïve Bayes rating scores (about 95 percent in average for each metric). Furthermore, despite introducing 10% noise, experimental findings reveal that the RF classifier performs better, although the Naïve Bayes classifier always performs worst. The results also indicate that the suggested voting

algorithm works better with respect to the precision, accuracy and reminder of measurements.

Total Number of Instances: 220111
Correctly Classified Instances: 220110 (99.9995 %)
Incorrectly Classified Instances: 1 (0.0005 %)
Mean absolute error: 0.0001
Relative absolute error: 0.0248 %
Root mean squared error: 0.0028
Root relative squared error: 0.8754 %
Kappa statistic: 1
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	0.971	0.000	1.000	0.971	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Fig. 6. Random Forest classifier (Noise) results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220111 (100 %)
Incorrectly Classified Instances: 0 (0 %)
Mean absolute error: 0
Relative absolute error: 0.0015 %
Root mean squared error: 0
Root relative squared error: 0.0012 %
Kappa statistic: 1
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Fig. 7. KNN classifier results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220090 (99.9905 %)
Incorrectly Classified Instances: 21 (0.0095 %)
Mean absolute error: 0
Relative absolute error: 0.0198 %
Root mean squared error: 0.0062
Root relative squared error:1.914 %
Kappa statistic: 0.9998
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	0.998	0.000	1.000	0.998	0.999	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	0.500	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Fig. 8. KNN Classifier (Noise) results summary

Total Number of Instances: 220111
Correctly Classified Instances: 209877 (95.3505 %)
Incorrectly Classified Instances: 10234 (4.6495 %)
Mean absolute error: 0.0171
Relative absolute error: 8.2129 %
Root mean squared error: 0.1177
Root relative squared error: 36.4597 %
Kappa statistic: 0.9104
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	0.920	0.019	0.976	0.920	0.994	**DoS**
	0.980	0.076	0.934	0.980	0.994	**DDoS**
	0.991	0.000	1.000	0.991	0.999	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	0.954	0.048	0.955	0.954	0.994	

Fig. 9. Naïve Bayes results summary

Table 2 compares the results of several ML algorithms on the UNSW-NB15 dataset. Experimental results show that the RF, KNN and Voting algorithms are the best performers with an accuracy of 100%, while Naïve Bayes classifier was the worst performer with an accuracy of 95.35%. Furthermore, after applying 10% of noise to the data, the experimental results indicate that the RF classifier has the highest performance. The results also reveal that the voting algorithm has the best performance over the evaluation metrics accuracy, precision, and recall.

Total Number of Instances: 220111
Correctly Classified Instances: 209792 (95.3119 %)
Incorrectly Classified Instances: 10319 (4.6881 %)
Mean absolute error: 0.0173
Relative absolute error: 8.2856 %
Root mean squared error: 0.1183
Root relative squared error: 36.6393 %
Kappa statistic: 0.9096
=== Detailed Accuracy by Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	0.919	0.019	0.976	0.919	0.994	**DoS**
	0.980	0.077	0.934	0.980	0.994	**DDoS**
	0.990	0.000	1.000	0.990	0.998	**Reconnaissance**
	0.941	0.000	1.000	0.941	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	0.953	0.049	0.954	0.953	0.997	

Fig. 10. Naïve Bayes (Noise) results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220111(100 %)
Incorrectly Classified Instances: 0 (0 %)
Mean absolute error: 0.0057
Relative absolute error: 2.7407 %
Root mean squared error: 0.0392
Root relative squared error: 12.1563 %
Kappa statistic: 1
=== Detailed Accuracy by Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Fig. 11. Voting Algorithm results summary

Total Number of Instances: 220111
Correctly Classified Instances: 220107 (99.9982 %)
Incorrectly Classified Instances: 4 (0.0018 %)
Mean absolute error: 0.0058
Relative absolute error: 2.7762 %
Root mean squared error: 0.0395
Root relative squared error: 12.2452 %
Kappa statistic: 1
=== Detailed Accuracy by Class ===

	TP Rate	FP Rate	Precision	Recall	ROC Area	Class
	1.000	0.000	1.000	1.000	1.000	**DoS**
	1.000	0.000	1.000	1.000	1.000	**DDoS**
	1.000	0.000	1.000	1.000	1.000	**Reconnaissance**
	1.000	0.000	1.000	1.000	1.000	**Normal**
	1.000	0.000	1.000	1.000	1.000	**Theft**
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	

Fig. 12. Voting Algorithm (Noise) results summary

Table 2
Comparison Results between the applied ML classifiers

ML algorithm	Accuracy	Precision	Recall
KNN	100	1	1
KNN (Noise)	99.9905	1	1
RF	100	1	1
RF (Noise)	99.9995	1	1
Naïve Bayes	95.3505	0.955	0.954
Naïve Bayes (Noise)	82.7719	0.820	0.828
Vote	100	1	1
Vote (Noise)	99.9982	1	1

5. Conclusions and Future Work

IoT networks have a variety of barriers to protection and privacy. This thesis has established a classification model to investigate the efficacy of renowned machine learning algorithms - namely RF, KNN and Naïve Bayes - in the

identification of attack signatures (i.e., abnormalities) of a dataset comprising both regular network traffic and malicious traffic instances. A series of assessment matrices, respectively accuracy, precision and reminder, were used to evaluate classifier efficiency. Both the KNN and RF classifiers displayed much greater efficiency than the Naïve Bayes classifier (in all measurement metrics). Moreover, a voting system that integrates many simple models was introduced in the proposed classificator model to boost the estimation outcomes and displays the better value over all measurement metrics. This research is confined to the results of three algorithms for machine learning to detect attack signatures only. In future, however, the work can be extended by observing the output of other learning classifiers using various assessment criteria on various data sets.

References

[1] Gupta, P.; Agrawal, D.; Chhabra, J.; Dhir, P.K. In Iot based smart healthcare kit, 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), 2016; IEEE: pp 237-242.

[2] Mustafa, G.; Ashraf, R.; Mirza, M.A.; Jamil, A. In A review of data security and cryptographic techniques in iot based devices, Proceedings of the 2nd International Conference on Future Networks and Distributed Systems, 2018; pp 1-9.

[3] Alaba, F.A.; Othman, M.; Hashem, I.A.T.; Alotaibi, F. Internet of things security: A survey. Journal of Network and Computer Applications 2017, 88, 10-28.

[4] Wang, J.; Chen, M.; Zhou, J.; Li, P. Data communication mechanism for greenhouse environment monitoring and control: An agent-based iot system. Information Processing in Agriculture 2019.

[5] Sicari, S.; Rizzardi, A.; Grieco, L.A.; Coen-Porisini, A. Security, privacy and trust in internet of things: The road ahead. Computer networks 2015, 76, 146-164.

[6] Xiao, L.; Wan, X.; Lu, X.; Zhang, Y.; Wu, D. Iot security techniques based on machine learning. arXiv preprint arXiv:1801.06275 2018, 1-20.

[7] Hameed, S.; Khan, F.I.; Hameed, B. Understanding security requirements and challenges in internet of things (IoT): A review. Journal of Computer Networks and Communications 2019, 2019, 1-14.

[8] Pishva, D. Iot: Their conveniences, security challenges and possible solutions. Adv. Sci. Technol. Eng. Syst. J 2017, 2, 1211-1217.

[9] Frustaci, M.; Pace, P.; Aloi, G.; Fortino, G. Evaluating critical security issues of the iot world: Present and future challenges. IEEE Internet of Things Journal 2017, 5, 2483-2495.

[10] Andročec, D.; Vrček, N. In Machine learning for the internet of things security: A systematic review, The 13th International Conference on Software Technologies, 2018.

[11] Cho, T.; Kim, H.; Yi, J.H. Security assessment of code obfuscation based on dynamic monitoring in android things. Ieee Access 2017, 5, 6361-6371.

[12] Ali, T.; Nauman, M.; Jan, S. Trust in iot: Dynamic remote attestation through efficient behavior capture. Cluster Computing 2018, 21, 409-421.

[13] Outchakoucht, A.; Hamza, E.; Leroy, J.P. Dynamic access control policy based on blockchain and machine learning for the internet of things. Int. J. Adv. Comput. Sci. Appl 2017, 8, 417-424.

[14] Wang, Z.; Chen, Y.; Patil, A.; Jayabalan, J.; Zhang, X.; Chang, C.-H.; Basu, A. Current mirror array: A novel circuit topology for combining physical unclonable function and machine learning. IEEE Transactions on Circuits and Systems I: Regular Papers 2017, 65, 1314-1326.

[15] Gebrie, M.T.; Abie, H. In Risk-based adaptive authentication for internet of things in smart home ehealth, Proceedings of the 11th European Conference on Software Architecture: Companion Proceedings, 2017; ACM: pp 102-108.

[16] Ahmed, M.E.; Kim, H.; Park, M. In Mitigating dns query-based ddos attacks with machine learning on software-defined networking, MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM), 2017; IEEE: pp 11-16.

[17] Li, Y.; Quevedo, D.E.; Dey, S.; Shi, L. Sinr-based dos attack on remote state estimation: A game-theoretic approach. IEEE Transactions on Control of Network Systems 2016, 4, 632-642.

[18] Tan, Z.; Jamdagni, A.; He, X.; Nanda, P.; Liu, R.P. A system for denial-of-service attack detection based on multivariate correlation analysis. IEEE transactions on parallel and distributed systems 2013, 25, 447-456.

[19] Razeghi, B.; Voloshynovskiy, S.; Kostadinov, D.; Taran, O. In Privacy preserving identification using sparse approximation with ambiguization, 2017 IEEE Workshop on Information Forensics and Security (WIFS), 2017; IEEE: pp 1-6.

[20] Yeh, K.-H.; Su, C.; Hsu, C.-L.; Chiu, W.; Hsueh, Y.-F. In Transparent authentication scheme with adaptive biometrie features for iot networks, 2016 IEEE 5th Global Conference on Consumer Electronics, 2016; IEEE: pp 1-2.

[21] Liu, J.; Zhang, C.; Fang, Y. Epic: A differential privacy framework to defend smart homes against internet traffic analysis. IEEE Internet of Things Journal 2018, 5, 1206-1217.