Computational Intelligence in IoT Attack Detection

Raju[1],Rajendra Kumar[1]

[1]*Department of Computer Science, Jamia Millia Islamia, New Delhi, INDIA*

Abstract: Computational Intelligence such as Artificial Intelligence, Machine Learning, Deep Learning and Fuzzy System are providing good opportunities to researchers to work on security of IoT devices. IoT devices are sensor embedded devices that interact with environment and collect data and this data is used to control these devices but if this data is not managed properly with the service providers then it may become a big cause of security breach. In this proposed paper Computational Intelligence technique Machine Learning KNN, Naive Bayes and Decision Tree were used to detect attacks in an IoT data set IoT_Fridge. For easy access various IoT datasets like Edge-IIoTset, TON-IoT, MQTT, Aposemat IoT, Bot-IoT, CTU-13 and MAWILab are also provided in this paper.Implementation and accuracy measurement of models performed on Jupyter Lab 2.1.5. Acurracy of KNN. Naive Baye and Decision Tree found 75%, 85% and 85% respectively. Finally, conclusion and future scope given at last of paper.

1. Introduction

Internet of Things that is commonly known as IoT, is not a new coin for people in present days. IoT is the fastest growing technology that is covering entire world and in coming few decades, it would be everywhere. Internet of Thing connects various electrical and electronic item with each other in such a way that they can communicate with each other by sending messages using a common platform. IoT enables users to take access of their belonging from remote areas. IoT also makes those tasks feasible that were not feasible before, for example through IoT device people can know about the weather of those places which were not in reach of people before. IoT is also play brilliant role in the field of industries, healthcare, education, home automation, energy saving and home care services.

There is no doubt that IoT is the best technology for coming generation but it is also true that it is in its infancy state where it has to face many problems and also face some criticism. The biggest challenge for IoT devices is Security. Platform through which IoT devices communicate with each others keep data at common place which can be Fog or Cloud and these platforms are handled by third party, since third party acts as a middle man between IoT and end users, hence there may be chances of breaching user security, therefore these platforms may be become cause of data leakage.It is true that many smart device manufacturing industries, researchers and scientists are working on security of Fog and Cloud but due to many complexities of IoT devices, they are not able to make fully and trustworthy IoT device for users.

In the last one decade the use of IoT devices increase exponentially that shows many people are preferring to use IoT devices because of its easy access. Today, around 50 billion IoT devices are in use and these devices are generating data in Zeta Byte (ZB) out of that 80% data is known as Dark IoT data which is not in use. This data is available to all at a very minimum cost or almost free. IoT has covered almost every field of modern life, it is in our home appliances, vehicles, offices, institutes, colleges and even in our handy handset that we always carry with us. Mobile phone generates our real time data over the internet and enables people to track us easily which is good for our known but it also breach our privacy. Yes, it is true that IoT is excellent technology that has to be opt by everyone but with this it is also important that we know the disadvantages of it.

Data is prime requirement of IoT devices and if we could secure this data then a secure IoT environment can be created. Data Security can be implemented by analyzing the pattern of data generated by the smart devices. If an IoT device monitor for a significant time then a normal use of device can be identified by seeing the data pattern. For example, if we monitor an IoT fridge for a period of six months and we found a normal pattern of data and suddenly we found some different types of data which is not similar to monitors data them there may be 90% chances of attack on the

device. In Data pattern recognition computational intelligence can play a very important role. Implementing Computational Intelligence based techniques like Artificial Intelligence, Machine Learning and Deep Learning can save IoT devices from unauthorized access. This paper propose machine learning algorithms to secure the IoT devices by analyzing collected data.

Machine learning algorithms like SVM, KNN, Naive bayes and Decision tree were used to detect different types of attacks in IoT data set IoT_Fridge. Output of each algorithm is compared with others and it was found that out of all these algorithms Naive bayes and decision Tree provide the 85% accuracy that is not too good but better to identify the attacks.

Further this paper organized as in Section-2 related work on IOT Security discussed, The proposed methodology explained in Section-3. To keep scarcity of IoT data and problem of finding good IoT data publicly available sources given in Section-4 with links. Section -5 focused on result and simulation and finally paper concluded with future scope in section-6.

2. Related work

This section of paper discuss the works of eminent researchers in the field of IoT security that is as follows:

In 2018, It was found that many IoT designers are facing various problems, especially the computational weakness of user-end devices [18] and it was also founded that attacks like DoS, M-in-M, Spoofing and other can be short-out by using various machine learning techniques.

A white list training model was proposed by Meidan et al. and he used it to predict unauthorized device on the network by using random forest [24]. Many predictive models are trained by Doshi eat al. by using several common machine learning algorithms and they founded a high degree of efficacy using this efficacy is referred as “stateless” features and stateless feature is defined as flow-independent characteristics of individual packets. In respect to secure IoT, many researchers have demonstrated that Machine Learning algorithms can be worked effectively to solve the security issues of IoT [29,31,32].

From the past few decades, researchers are working to prove that machine leaning techniques are good to secure IoT systems. In 2007, Moskovitch et. al. proposed a model for prediction of worm activity with 90% accuracy and in 2013, awarded a patent to file a system for detecting malicious behavioral pattern in computer using machine learning [34]. A telemetry-based machine learning approach was proposed by Ponomarev and Atkison and they also demonstrated accuracy more than 90% for some model [27].

In IoT, Edge devices are more vulnerable to attack because physical availability of these devices might be the greatest opportunity for attackers to steal sensitive and important information. While IoT system follow the heterogeneity in nature but still there are some homogeneous constraints like Internet Connectivity through which IoT devices communicate with each other and this common platform also give opportunities to attackers to attack over the system. The globally availability of volumetric IoT data is also big challenge for IoT security and to stop this global availability of IoT data, the concept of “Fog” is going to use and this concept not only stop the globally availability of volumetric IoT data but also improve the security of system [23], but with this advantage Fog computing also have some disadvantages like complexity that increase the authentication problem, securing transient data and maintaining user sensitive information [28].

Till now very few research paper published on Computational Intelligence based security for IoT and very few paper worked on ML/DL to secure IoT devices. In paper [1] a survey of IoT communication on security issues with solution is discussed. [2], emphasized on Intrusion detection for IoT systems.

TABLE 1: Literature Review

SL.No	Authors	Year	Contribution	Ref.
1.	Xiao et al.	2018	Machine Learning approaches discussed to understand the attacks on IoT like DoS, Man-in-Middle and spoofing and they also conclude that ML algorithms can be used to solve these problems.	[29]

2.	Meidan et al.	2017	A whitest training model was proposed, and random forest is used to predict unauthorized device on network.	[24]
3.	Doshi et al.	2018	Machine Learning algorithms based predictive models were evaluated to find a high degree of efficacy.	[25]
4.	Moskovitch et al.	2007	A demonstration is presented to predict worm activity.	[33]
5.	Moskovitch et al.	2013	Awarded a patent to detect malicious behavioural patterns in computer using ML.	[34]
6.	Ponomarev and Atkison	2015	A telemetry-based machine learning approach was proposed and demonstrated accuracy for some models more than 90%	[27]
7.	Mahmood et al.	2019	Fog computing concept was proposed to keep voluminous IoT data local and improve the security.	[23]
8.	Hassija et al.	2019	Complexity in authentication, securing transient information and maintaining user privacy due to Fog computing are discussed	[28]
9.	Pa et al.	2016	A model was designed to investigate on going attacks against Telnet services and discovered several malware families.	[7]
10.	Guarnizo et al.	2017	The Scalable High-Interaction Honeypot (SIPHON) was proposed for the IoT paradigm and demonstrated how the combination of a limited number of physical devices and worldwide wormholes permit the emulation of numerous IoT devices on the Internet.	[8]
11.	Costin et al.	2014	A large-scale static analysis of embedded firmware to explore IoT insecurities is performed.	[12]
12.	Meidan et al.	2017	Classified IoT nodes connected to an organization's network by solely observing network traffic.	[15]
13.	Meng et al.	2018	Various security solutions to secure IoT devices are discussed after analysing various challenges like spoofing, jamming and unauthorized access.	[5]
14.	Hassan et al.	2019	Current trends in IoT Security are studied.	[7]
15.	J. Granjal, et al.	2015	Security issues with solution for IoT Communication	[1]

3. Proposed Methodology

This research proposed various Supervised Machine Learning Algorithm to secure our data received through IoT device. Collected data will be classified into two main categories Secure Data (D_{secure})and Attack Data (D_{attack}) and whenever device receive any New Data (D_{new}), our research compare distance of D_{attack} with both the cluster (D_{secure}&D_{attack}) and categorize the D_{new} into D_{secure} or D_{attack}.

Algorithm

1. Start
2. Choose IoT Data Set (D_{iot})
3. Apply Knowledge Discovery Database (KDD) steps on D_{iot}
4. Classify D_{iot} into D_{secure} and D_{attack} by using KNN, Naïve Bayes and Decision Tree

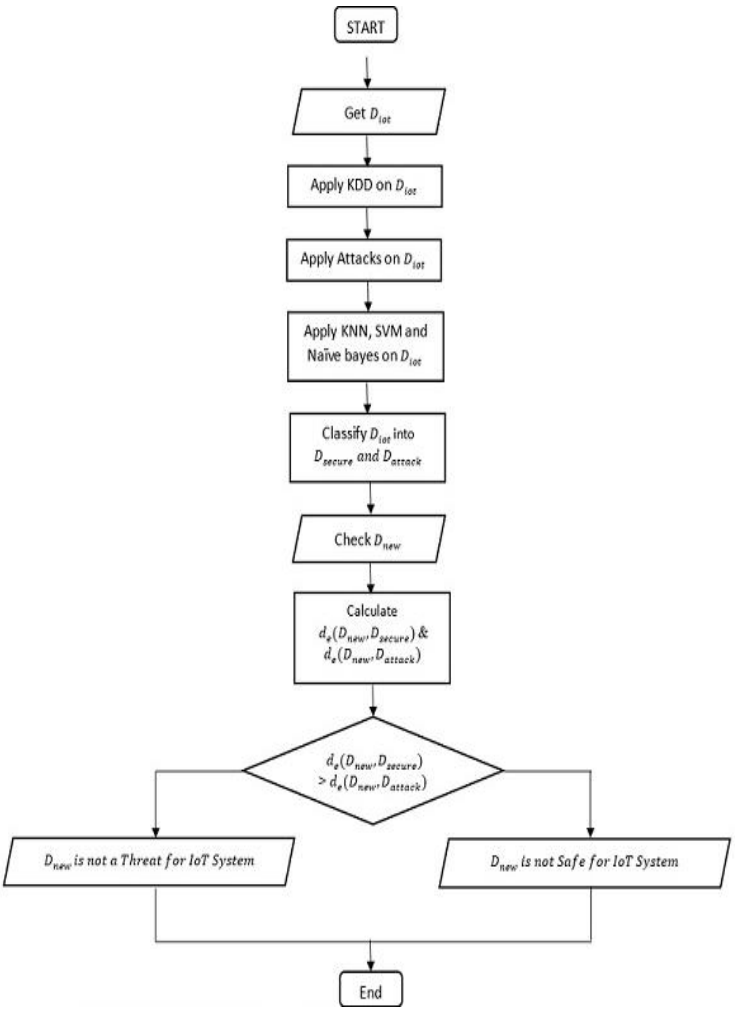


Figure 1. Flow Chart of Proposed Work

The purpose of this article is to secure IoT devices using data collected by them. As security is provided by analyzing the data therefore this security is known as Data Driven Security. In smart world the volume of data is increasing with very fast speed and this growth of data also providing opportunities to researchers to improve the security of smart devices. This paper will focus on Machine Learning [3] & [4], Data mining [5]-[8], Data Visualization [8], [9] and Data Analytic [10] techniques to provide Data Driven Security to IoT device.

3.1 Computational Intelligence

Now the time has come when machines are ready to mimic human activities and now artificial neurons are taking the place of biological neurons. The technology that is enabling machines to think like human is termed as Computational Intelligence (CI) and it has three main stumps Neural Networks, Fuzzy Systems and Evolutionary Systems.

Artificial Intelligence, Database Management System (DBMS) and Decision Support System (DSS)enhance the impact of CI in several engineering applications [14]. CI techniques like Machine Learning and Deep Learning can be used to secure the IoT devices by managing streaming as well as big data received via IoT devices. The Supervised Algorithms are classified intoSupport Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbors (KNN),Ensemble Learning, Decision Tree (DT) and Association Rule (AR) while Unsupervised Learning consists K-mean and Principal Component Analysis (PCA). Deep Learning Techniques is alsodivided into Supervised, Unsupervised and hybrid methods. In this paperdiscuses few machine learning algorithms that would be used in security of IoT.

Machine Learning is not a new coin for the data scientist and researchers; they all are familiar with this term and using this technology in prediction and securing data from unauthorized access. Machine Learning is categorized into three types learning that are Supervise Learning, Unsupervised Learning and Reinforcement Learning. In Supervise Learning, machine has to be trained on some label data and need some model to map input data with output data. In simple language, we can say that supervised learning is just as class learning as we have teacher in class to train students; similarly we have output to map input data. Mathematically, it is represented as

$$O_p = f(I_p) \quad (1)$$

Where O_p *is output data and* I_p *is input data.*

5. Comparing D_{new} with D_{secure} *and* D_{attack}
6. Calculating $d_e(D_{new}, D_{secure})$ and $d_e(D_{new}, D_{attack})$
7. If $d_e(D_{new}, D_{secure}) > d_e(D_{new}, D_{attack})$
8. Print " D_{new} have some Security Issues"
9. Else:
10. Print " D_{new} is Safe from Hackers"
11. End

The above algorithm is shown in flow chart 1 that shows the flow of our work.

Machine Learning is going to use in most of the fields of real life. It is used in medical for disease prediction, weather forecasting, departmental stores, digital marketing etc. Few more real-world applications of Machine Learning are as follows:

- Character Recognition in security encryption in different handwriting style.
- Face recognition in forensic
- Malicious Code identification in Software and apps.

In this paper, few supervised algorithms are use to secure data received by a smart device.

3.2 Supervised Learning Algorithms

This section briefly described the various supervised machine learning algorithms Naive Bayes, SVM, k-NN and Decision and explain how these algorithm implemeted to secure IoT devices by analyzing data of IoT devices.

3.2.1. Naïve Bayes (NB)

NB is a supervised learning algorithm based on the principal of conditional independence probability. In this supervised learning algorithm, existence of one event is totally independence from other event; hence each event has equal effect on the result. Multinomial, Bernoulli and Gaussian are three type of Naïve Bayes classifier. [15]

Mathematically, Naïve Bayes algorithm probability represented as follows

$$P(F/G) = \frac{P(F \cap G)}{P(G)}$$

(2)

Where,
F *is occuring feature and* G *is Given feature,* $P(F/G)$ *is Probability of* F *given that* G $P(F \cap G)$ *is probability of occuring feature* F *and* G.

3.2.2. Support Vector Machine (SVM)

Another supervised machine learning algorithm is Support Vector Machine (SVM) that is used to classify data on the basis of some features. SVM first of all plot the data points on a space of n-dimensional and draw a hyperplane which classify the data into two categories perfectly.[16]

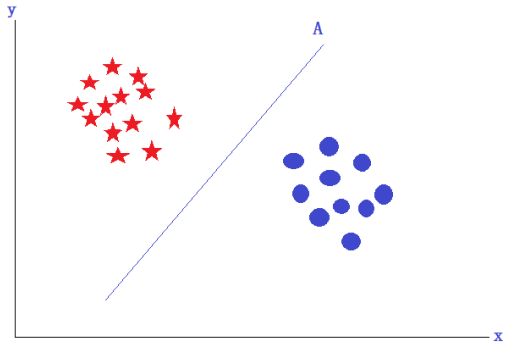


Figure. 2. Hyperplane A, classifying both classes

In SVM, Hyper plane play a very important role to classify the data but main problem arises when we have more than one hyper plane, in case of more than one hyper plane following rules are used to choose right hyperplane.

Rule 1: A right hyperplane always sperate two classes perfectly. In Fig. 2, we have three hyper-planes A, B and C but Plane B is the right hyperplane because it is perfectly classifying both classes.

Rule 2: In this case when more than one hyperplane are classifying classes perfectly, then we have to choose that hyperplane which has maximum distance from the nearest data points of both the classes.

For Example, in Fig. 3 C is the best hyper plane to separate two classes because it has maximum distance from the both classes nearest data points.

Rule 3: In Fig. 4, hyper plane A is classifying both classes as discuss in Rule1 and Rule 2

Rule 4: In some cases, hyper planes are not able to classify the classes because few different class data lies in other class such that they can't be separated by any hyper plane. In that case we found Outlier which is found in that class through which it doesn't belong. In Fig. 5, we can see hyper plane is not able to classify both the classes perfectly, in this type of case we get some outliers that may be big threat to the security of system.

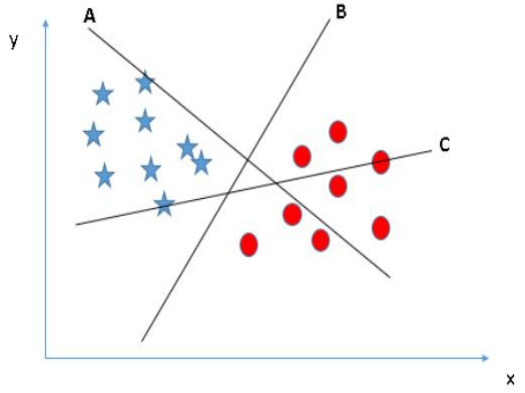


Figure. 3. Hyper plane B, classifying both classes

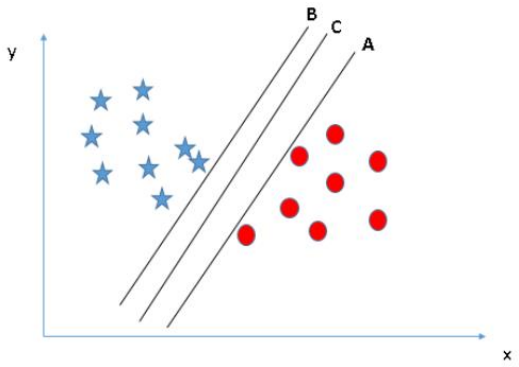


Figure. 4. Hyper plane C, classifying both classes

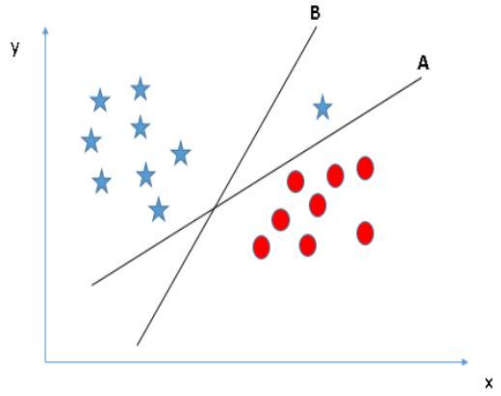


Figure5. Hyper plane A, classifying both classes

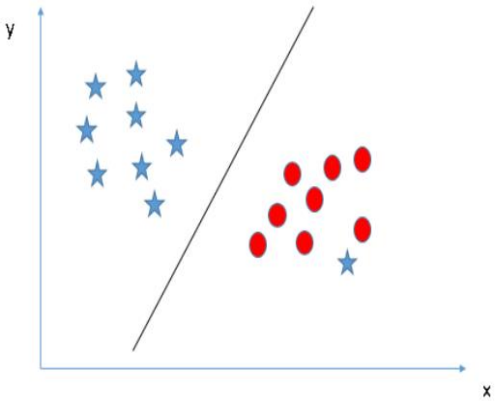


Figure. 6. Hyper plane not able to classify (Outlier)

3.2.3. K-Nearest Neighbor (KNN)

K-Nearest Neighbor is another supervised machine learning algorithm that is useful for predictive problem after classifying data. KNN works on feature matching, in this approach K clusters are formed on the basis of similar features of datum and the distance of new data point is calculated from the mean of each cluster. Following Distance Formulae are used to calculate distance

$$1.\ Euclidian\ Distance, d_e = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

$$2.\ Manhattan\ Distance, d_m = \sum_{i=1}^{n}|x_i - y_i| \quad (4)$$

$$3.\,Hamming\ Distance, d_h(x.P, y.P) = \begin{cases} 0 & if\ x.P = y.P \\ 1 & if\ x.P \neq y.P \end{cases} \quad (5)$$

In KNN, Euclidian Distance (Eq. 4) is mostly used. After calculated distance of new data point sent to that cluster from where it's Distance is minimum.

3.2.4. Decision Tree (DT)

DT is an important Supervise Machine Learning algorithm that is used in both classification and regression. It is a non-parametric algorithm with tree like structure which consist root node, branches, leaf nodes and intermediate nodes. Root node is the node in decision tree which has no incoming edge and it is the beginning of Decision Tree. Root node is selected by using Features Selection Methods (FSM), FSM also helps to reduce the features of available data set which make our analysis better and precise. Gini Index and Information gain are two important methods of selecting root node.

3.2.4.1 Information Gain (IG)

In root node selection, information gain plays a very important role. It provides the node/feature which has highest information about the class. Impurity, uncertainty and entropy are the concept on which Information gain is based, IG tells about the randomness of features in a class, the highest randomness disqualify the node for the root node selection. IG also reduces the entropy of DT from originate node to terminal nodes. It varies from 0 to 1, where 0 and 1 represent same level of entropy which mean low entropy because all objects belong to one class while 0.5 shows highest level of entropy where objects of all classes equally distributed [22]. Entropy is mathematically expressed as follows (eq. 5)

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i \quad (6)$$

Information Gain is calculated using following formula

IG = Entropy (Parent Node) – Average Entropy (Child Nodes) (7)

To illustrate Information Gain Calculation, refer the following example

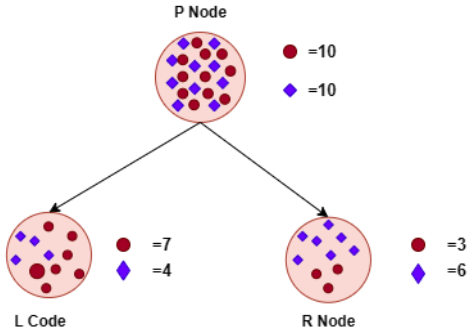


Fig. 7. Example of Decision Tree Root Node selection

$$E(P) = -\frac{10}{20}\log_2\frac{10}{20} - \frac{10}{20}\log_2\frac{10}{20} = 1.0$$

$$E(L) = -\frac{7}{11}\log_2\frac{7}{11} - \frac{4}{11}\log_2\frac{4}{11} = 0.946$$

$$E(L) = -\frac{3}{9}\log_2\frac{3}{9} - \frac{6}{9}\log_2\frac{6}{9} = 0.918$$

$$Weigted\ Average\ Entropy\ (L,R) = \frac{11}{20} \times 0.946 + \frac{9}{20} \times 0.918 = 0.933$$

$$Information\ gain = 1.0 - 0.993 = 0.067$$

In this example, information gain is very low which indicates that this node is not suitable for Root Node.

3.2.4.2 Gini Index or Gini Impurity

Gini Index is a good way to measure the entropy of variables from originate node to terminal nodes in a DT. Degree of Gini index varies in [0,1], where 0 represents no error/entropy in variable that means all variables belong to one class and 1 represents high impurity in variable which means all variables belongs to different class. If Gini Index is equal to 0.5 then it is considered as all variables are equally distributed in available all classes. So, lower value of Gini Index is preferable in Decision Tree. Gini Index is calculated by using the following equation (Eq. 5) [22]

$$GI = 1 - \sum_{i=1}^{n} (p_i)^2 \quad (8)$$

Where, GI = Gini Index/ Gini Impurity, p_i = probability of an element being send to one particular class**.**

4. **Publicly available IoT Data Sets**

This section focus on the openly available IoT data sets because most of the time IoT data set become big hindrance in deployment of security of devices. This section briefly discuss the IoT data sets with available attacks. In Table 2, available datasets are shown with link [30]

TABLE 2. IoT Data Set

IoT Data set	Year	Free	Attack	Format	Realistic	Link
Edge-IIoTset [30]	2022	YES	DoS/DDoS attacks,Information gathering, Man in the middle attacks, Injection attacks, Malware attacks	Packets, Tabular	YES	https://www.kaggle.com/datasets/mohamedamineferrag/edgeiiotset-cyber-security-dataset-of-iot-iiot
TON-IoT [31]	2020	YES	backdoor,ddos,dos,injection, mitm, password, ransomware, scanning, xss	Packets, Tabular	YES	https://cloudstor.aarnet.edu.au/plus/s/ds5zW91vdgjEj9i
MQTTTest [32]	2020	YES	SlowITe,Bruteforce,Malformed data, Flooding,DoS attack	Packets, Tabular	YES	https://www.kaggle.com/datasets/cnrieiit/mqttset
Aposemat IoT-23[33]	2020	YES	Botnet, C2, Background	Packets, Flow	YES	https://www.stratosphereips.org/datasets-iot23
Bot-IoT[34]	2019	YES	DDoS, DoS, OS and Service Scan, Keylogging,Data exfiltration	Packets, Tabular	YES	https://cloudstor.aarnet.edu.au/plus/s/umT99TnxvbpkkoE
N-BaIoT [35]	2018	Yes	Bashlite,Mirai	Tabular	YES	https://www.kaggle.com/datasets/mkashifn/nbaiot-dataset
CTU-13 [36]	2014	Yes	Botnet, C2, Background	Packet, Malware	Yes	https://www.stratosphereips.org/datasets-ctu13

MAWILab [37]	2010	No	Attack, Special,Unknown	Tabular	Yes	http://www.fukuda-lab.org/mawilab/data.html

5. Results and Simulation

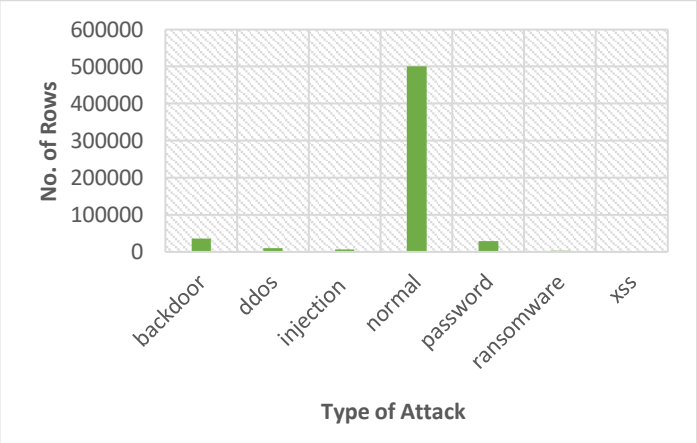
This section explained the data set used for experiment and simulation for proposed work. Data set was closely analyse to get it ready for experiment, required preprocessing steps are implemented over the data set to avoid any kind of anomaly due data set. Performance metrices are also described to compare result of proposed models. Three machine learning algorithms Naive Bayes, KNN and Decision tree are applied over data set to detect attacks.

5.1 Data set

IoT_Fridge data set is used for proposed model this data set contains 587077 rows and 5 columns that has 500827 normal data and 86,249 attack data which is suffered from various attacks as shown in Table 3. and same data is shown through bar chart.

Table 3. Attack wise Data distribution

Type of Attack	No. of rows
backdoor	35568
ddos	10233
injection	7079
normal	500827
password	28425
ransomware	2902
xss	2042



5.2 Proposed Methodology

To apply machine learning algorithms over data, first of all data set (IoT_Fridge) is collected and deeply analyzed and after analyzing required Knowledge Discovery in Databases (KDD) steps are applied to get clean data. Finally, 6 types of attacks found in data set that are as follows:

DDoS Attack: Distributed Denial of Service (DDoS) attack is the most challenging work for IoT Security researchers. DDoS attackers use the disadvantage of IoT devices like limited Storage Capacity and limited Network Capacity to get the access of IoT device and make it unavailable for the authorized users.[23]

Injection Attack: Injection attack is now become a very common and easy for hackers. In Injection attack, attackers can get the access of IoT device without a valid password and inject the malicious code or instruction with the valid instructions that become a cause of stealing data.[23]

Backdoor Attack: IoT devices manufactures creates some hidden access methods for accessing IoT devices to support their users and these hidden methods are known as backdoors than can be UserId or Password, but these backdoors act as a front door for the hackers and provide them good opportunities to get access of IoT devices and make some money by blackmailing the users. [20]

XSS: Cross-site Scripting is known known as xxs, it is a web based vulnerability for IoT devices that can add a dangerous code to the IoT devices without knowing the authentic user.[21]

Password: Password is the most common credential that known by most of the users but only few users knows the features of strong password and most of the people use very simple password and even the default password to protect their devices that becomes a good chance for hackers to steal their data. So, without a proper management of password, IoT Security can be sustained in the market.
Ransomeware: in this attack, attackers encrypt the file due to which authentic user not able to access that file and then attacker sell the decryption key to user.

5.3 Hardware and Software Requirement

5.3.1 Device specifications

Device name	DESKTOP-603UK47
Processor	Intel(R) Xeon(R) Silver 4208 CPU @ 2.10GHz 2.10 GHz
Installed RAM	64.0 GB (63.6 GB usable)
Device ID	63AB20DA-4805-4A4B-89CF-20F41FA015ED
Product ID	00330-71462-46424-AAOEM
System type	64-bit operating system, x64-based processor
Pen and touch	No pen or touch input is available for this display

5.3.2 Window Specification

Edition	Windows 10 Pro
Version	22H2
Installed on	22-03-2022
OS build	19045.3086
Experience	Windows Feature Experience Pack 1000.19041.1000.0
Tool	ANACONDA NAVIGATOR Jupyter Lab 2.1.5

5.4 Algorithm

1. load data set IoT_Fridge.

2. Apply required KDD steps

3. Identify Target value (‘type’)

4. Split data set in Test and Train data set (Train = 80% and Test 20%)

5. Scaling the values

6. Create KNN, Naive Bayes and Decision Tree model and apply on data set one by one.

7. Perform Accuracy Measurement

8. Predict the attack.

5.5 Accuracy Measures for Machine Learning Algorithms

Accuracy Measurement is the main task to know the effectiveness of applying computational intelligence techniques. As we use KNN, SVM and Naïve Bayes algorithms to secure our IoT System. So, to measure their accuracy, we will use the following evaluation matrices.[17]

1. Confusion Matrix
2. Area under Curve
3. F1-Score

5.5.1. Confusion Matrix

Confusion Matrix provides us a matrix after analyzing data that contains output and complete performance of model. In general confusion matrix is represented as shown in Fig.. 7. In Confusion Matrix four important terms are used which are

True Positive (TP): When **Predictive True** value is same as the **Actual True**

True Negative (TN): When **Predictive Negative** value equal to **Actual Negative** value

False Positive (FP): When **Predictive value is TRUE** but **Actual value is FALSE**

False Negative (FN): When **Predictive value is FALSE** but **Actual value is TRUE**

All these parameters of confusion matrix help to find out the following measurement of a model.

Accuracy: Accuracy of the model is the ratio of Truly Predictive all values to Total of all values, as shown in Equation 5.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \quad (8)$$

5.5.2. Area under Curve (AUC)

Area under Curve is the curve that is drawn between True Positive Rate (TPR) and False Positive Rate (FPR) at different data points. It measures the ability of classifier to classify two classes perfectly. To know AUC, it is compulsory that we make ourselves familiar with the following terms.

True Positive Rate/Sensitivity/Recall: TPR represents the correctly predicted data points. It is calculated by using following equation (Eq. 6).

$$\boldsymbol{Sensitivity} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \quad (9)$$

True Negative Rate/ Specificity: Specificity gives the proportion of Negative Class that is correctly predicted. Eq. 7 is used to calculate it.

$$\boldsymbol{Specificity} = \frac{\boldsymbol{TN}}{\boldsymbol{FP+TN}} \quad (10)$$

False Positive Rate: It represents the proportion of Negative class that is incorrectly classified, formula to calculate it is shown in Eq. 8

$$\boldsymbol{FPR} = \mathbf{1} - \boldsymbol{Specificity} = \mathbf{1} - \frac{\boldsymbol{TN}}{\boldsymbol{FP+TN}} = \frac{\boldsymbol{FP}}{\boldsymbol{FP+TN}} \quad (11)$$

Different values of AUC represent the different level of classification. Higher AUC represents higher accuracy in classification and lower represent lower accuracy. Table 1 show different range of AUC for classification.

Table 3. AUC measurement range

AUC	Level of Accuracy
0	Prediction is totally opposite from the correct data. It means it is representing Positive Values as Negative and Negative Values as Positive
1	Perfectly distinguish Positive and Negative values. It means classifier representing Positive Values as Positive and Negative Values as Negative
0.5	If AUC=0.5, the AUC is not able to distinguish between Positive & Negative Values.
0.5<AUC<1.0	This range of AUC represents high chance of getting correct prediction.

5.5.3. F1-Score

F1 score is the harmonic mean of Precision and Recall, and it is calculated to know the balance between Precision and Recall. F1-Score also represents the relation between Precision and Recall as represent in eg. 12

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

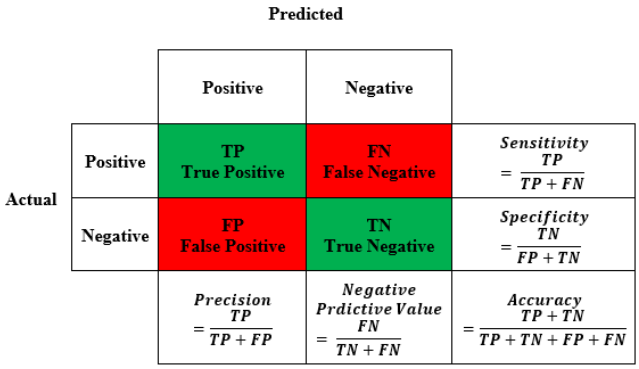


Figure 8. Confusion Matrix

5.6 Training and Classifier

In this proposed work KNN, Naive Bayes and Decision Tree machine learning algorithms are used to detect attacks. Accuracy, Precision, Recall and F1-Score of these three models are shown in Table 4.

Table 4. Accuracy Comparison of k-NN, Naïve Bayes and Decision Tree

Algorithm	**k-NN**	**Naïve Bayes**	**Decision Tree**
Accuracy	0.78	0.85	0.85
Precision	0.91	0.85	0.85
Recall	0.89	1	1
F1-Score	0.90	0.92	0.92

Out of these three models it was found that both Naive Bayes and Decision Tree are giving the same result 85% accuracy while K-NN providing low accuracy 78%.

6. Conclusion and Future Scope

This research focused to detect the attacks in IoT data set, in this paper IoT_Fridge data set is used in which six types of attacks given with normal data. To detect attacks, computational intelligence techniques like machine learning is used. KNN, Naive Bayes and Decision Tree used to detect attacks and then accuracy is found of all these three algorithms. Accuracy of Naive Bayes and Decision Tree found same that 85% which k-NN gave 78%. This proposed work also encouraged the young researchers to improve the accuracy of these models by applying other computational intelligence techniques like Artificial Intelligence, Deep Learning amd Evolutionary computation.

REFERENCES

[1]. J. Granjal, E. Monteiro, and J. Sa Silva, “Security for the internet of things: a survey of existing protocols and open research issues,” IEEE Communications Surveys & Tutorials, vol. 17, no. 3, pp. 1294–1312, 2015.

[2]. B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, “A survey of intrusion detection in Internet of)ings,” Journal of Network and Computer Applications, vol. 84, pp. 25–37, 2017.

[3]. N. Shone, T. Nguyen Ngoc, V. Dinh Phai, and Q. Shi, ‘‘A deep learning approach to network intrusion detection,’’ IEEE Trans. Emerg. Topics Comput. Intell., vol. 2, no. 1, pp. 41–50, Feb. 2018.

[4]. A. Javaid, Q. Niyaz, W. Sun, and M. Alam, ''A deep learning approach for network intrusion detection system,'' in Proc. 9th EAI Int. Conf. Bio-Inspired Inf. Commun. Technol. (BIONETICS), 2016, pp. 21–26.
[5]. R. Sahani, C. Rout, J. C. Badajena, A. K. Jena, and H. Das, ''Classification of intrusion detection using data mining techniques,'' in Progress in Computing, Analytics and Networking. Singapore: Springer, 2018, pp. 753–764.
[6]. P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava, and P.-N. Tan, ''Data mining for network intrusion detection,'' in Proc. NSF Workshop Next Gener. Data Mining, 2002, pp. 21–30.
[7]. L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, ''Information security in big data: Privacy and data mining,'' IEEE Access, vol. 2, pp. 1149–1176, 2014.
[8]. M. Kolomeec, A. Chechulin, A. Pronoza, and I. V. Kotenko, ''Technique of data visualization: Example of network topology display for security monitoring,'' JoWUA, vol. 7, no. 1, pp. 58–78, 2016.
[9]. H. Shiravi, A. Shiravi, and A. A. Ghorbani, ''A survey of visualization systems for network security,'' IEEE Trans. Vis. Comput. Graphics, vol. 18, no. 8, pp. 1313–1329, Aug. 2012.
[10]. A. Cárdenas, P. K. Manadhata, and S. P. Rajan, ''Big data analytics for security,'' IEEE Security Privacy, vol. 11, no. 6, pp. 74–76, Nov./Dec. 2013.
[11]. D. Han, Y. Mo, and L. Xie, ''Convex optimization based state estimation against sparse integrity attacks,'' IEEE Trans. Autom. Control, vol. 64, no. 6, pp. 2383–2395, Jun. 2019.
[12]. B. Mao, Y. Kawamoto, and N. Kato, ''AI-based joint optimization of QoS and security for 6G energy harvesting Internet of Things,'' IEEE Internet Things J., early access, Mar. 23, 2020, doi: 10.1109/JIOT.2020.2982417.
[13]. T. Godquin, M. Barbier, C. Gaber, J.-L. Grimault, and J.-M.-L. Bars, ''Placement optimization of IoT security solutions for edge computing based on graph theory,'' in Proc. IEEE 38th Int. Perform. Comput. Commun. Conf. (IPCCC), Oct. 2019, pp. 1–7.
[14]. Access from: https://www.routledge.com/Computational-Intelligence-in-Engineering-Problem-Solving/book-series/CIEPS
[15]. Access from: https://www.ibm.com/cloud/learn/supervised-learning
[16]. Access from: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234
[17]. Access from: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234
[18]. Access from: DDoS attacks in IoT networks: a comprehensive systematic literature review | SpringerLink
[19]. Access from: Understanding IoT Vulnerabilities: SQL injection or Hackers Can Hit Connected Things with Tricky Requests (bitdefender.com)
[20]. Access from : Trending IoT Malware Attack | How To Protect Your IoT Devices (biz4intellia.com)
[21]. Access From: Understanding IoT Vulnerabilities: Code Injection Attacks Can Steal Your Web Life (bitdefender.com)
[22]. Access from: Gini Index: Decision Tree, Formula, and Coefficient (quantinsti.com)
[23]. Mahmood, S.; Ullah, A.; Kayani, A.K. Fog Computing Trust based Architecture for Internet of Things Devices. Int. J. Comput. Commun. Networks 2019, 1, 18–25.
[24]. Mahmood, S.; Ullah, A.; Kayani, A.K. Fog Computing Trust based Architecture for Internet of Things Devices. Int. J. Comput. Commun. Networks 2019, 1, 18–25.
[25]. Doshi, R.; Apthorpe, N.; Feamster, N. Machine learning ddos detection for consumer internet of things devices. In Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW), Francisco, CA, USA, 24 May 2018; pp. 29–35
[26]. Miettinen, M.; Marchal, S.; Hafeez, I.; Asokan, N.; Sadeghi, A.R.; Tarkoma, S. IoT Sentinel: Automated device-type identification for security enforcement in IoT. In Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, USA, 5–8 June 2017; pp. 2177–2184.
[27]. Ponomarev, S.; Atkison, T. Industrial control system network intrusion detection by telemetry analysis. IEEE Trans. Dependable Secur. Comput. 2015, 13, 252–260. [CrossRef]
[28]. Hassija, V.; Chamola, V.; Saxena, V.; Jain, D.; Goyal, P.; Sikdar, B. A survey on IoT security: Application areas, security threats, and solution architectures. IEEE Access 2019, 7, 82721–82743. [CrossRef]
[29]. Xiao, L.; Wan, X.; Lu, X.; Zhang, Y.; Wu, D. IoT security techniques based on machine learning. arXiv 2018, arXiv:1801.06275

[30]. MohamedAmine Ferrag,Othmane Friha,Djallel Hamouda,LeandrosMaglaras, andHelge Janicke. 2022. Edge-IIoTset:ANew Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications: Centralized and Federated Learning. (2022). https: //doi.org/10.21227/mbc1-1h68.

[31]. 2021. A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets. Sustainable Cities and Society 72 (2021), 102994. https://doi.org/10.1016/j.scs.2021.102994.

[32]. SalmanRachmadi,SatriaMandala,andDitaOktaria.2021. DetectionofDoSAttackusingAdaBoostAlgorithmonIoTSystem.In2021 International Conference on Data Science and Its Applications (ICoDSA). IEEE, 28–33.

[33]. AgustinParmisano SebastianGarcia andMaria JoseErquiaga.2020. IoT-23:Alabeleddataset withmalicious andbenignIoTnetwork traffic (Version1.0.0)[Data set]. (2020). https://doi.org/10.21227/mbc1-1h68

[34]. Nickolaos Koroniotis, Nour Moustafa, Elena Sitnikova, and Benjamin Turnbull. 2019. Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. Future Generation Computer Systems 100 (2019), 779–796. https://doi.org/10.1016/j.future.2019.05.041

[35]. YairMeidan,MichaelBohadana,YaelMathov,YisroelMirsky,DominikBreitenbacher,AsafShabtai, andYuvalElovici. 2018. NBaIoT: Network-based Detection of IoT Botnet Attacks Using Deep Autoencoders. IEEE Pervasive Computing 17, 3 (July 2018), 12–22. https://doi.org/10.1109/MPRV.2018.03367731 arXiv: 1805.03409.

[36]. S. García, M. Grill, J. Stiborek, and A. Zunino. 2014. An empirical comparison of botnet detection methods. Computers & Security 45 (Sept. 2014), 100–123. https://doi.org/10.1016/j.cose.2014.05.011

[37]. [n.d.]. MAWILab.([n.d.]). http://www.fukuda-lab.org/mawilab/