Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 6, 9702 – 9712

Evaluation of Machine Learning Algorithm using Hybrid classifier for IDS in Infrastructure less wireless networks

[1]Sheela Devi, [2]R.K. Chauhan, [3]Ashwani Kush

[1]Research Scholar, Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, Haryana, India, sheela.sharma.kkr@gmail.com
[2]Professor, Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, Haryana, India, rkchauhandcsakuk@gmail.com
[3]Associate Professor, Department of Computer Science, IIHS, Kurukshetra University, Kurukshetra, Haryana, India, akush20@gmail.com

Abstract

Intrusion detection system (IDS) can be hardware or software which is used in networking to monitor the network. This is useful when malicious activity is to be managed for network however mounting an IDS is challenging task because attackers always find a different way to attack in the network. Several research has been done on IDS but still there are some issues exists like de An abundance of similar researches had done to related areas which used machine learning process into both host and network-based intrusion systems. But doing the same initially presented few problems like accuracy, high false alarm rate ,low detection rate etc. nowadays due to the very much development in of machine learning which outcome the vast improvements in machine learning algorithms . This paper present a framework model in which many classifiers are used to detect the malicious activity by using machine learning methods. Research uses historical data to apply Intrusion detection on Host machine to detect malicious node. NSL-KDD datasets are used to test the data and evaluation. Based on the NSL-KDD datasets base station is trained so that if any malicious activity is done on the machine then IDS will detect that malicious node based on anomalous behaviour. We define abnormal behaviour by deviating from their normal behaviour. For the training of Host machine Supervised Learning is used which requires already been labelled training data and supervised learning algorithms are used for prediction. Our research also used some data mining algorithms such as KNN (k-nearest neighbour), decision trees, and Ensemble and hybrid algorithms. Hybrid algorithm is the combination of KNN, DT, and Ensemble Algorithm which is more reliable in terms of performance metrics. Results based on evaluation also demonstrate that the hybrid algorithm is best one among the other classifiers which are used in this research and hybrid algorithm attained the highest accuracy (98.99) and the lowest false positive rate (1.01).

Keywords: Host machine, Supervised Learning, Intrusion Detection, data mining algorithms; NSL-KDD.

1. Introduction

Mobile computing has become trends and now become necessity in daily life due to the digital revolution over worldwide in mobile technology and mobile devices. Today's people are using palmtop i: e mobile phones, PDAs or a laptop in which the these devices or nodes are not bound to any centralized control like base stations or mobile switching centres for sharing information and their daily activities like reading newspaper, students download study materials, audio and video clips etc.so people

are accessing wireless networks for doing the same at any time.

Wireless networks are categorized into two class

A. Infrastructured and

B. Infrastructure less networks.

Infrastructured wireless networks: - These type of networks are bound to fixed structures and all the stations or nodes are restricted that they have to tie with base station through using wires. This type of network have wireless access points i:e routers and all the devices like computers,mobile,Tablets,printers etc. are connected wirelessly to this access point .Cellular phone network is the example of infrastructure wireless networks.

Infrastructure less networks:-There is no any fixed infrastructure is required for this type of network.so no access point (I: e routers) or no access control is require. Nodes in this type of networks can roaming and communicate to their devices and connection of nodes can be done dynamically. In this way the whole network becomes mobile. Mobile Ad hoc Networks (MANETs) is the example of this category. Due to the mobility of nodes information of routing table will be changing according to their location.so path of nodes will also be change as per instant of time.

Intrusion detection system plays a very imp. Role in detecting attacks.

There are three methods of detecting attacks using IDS: signature based detection method, anomaly based detection method and hybrid-based detection method.

Signature-based detection use already known instruction sequence (sequence of 1's and 0's) and it fails when new pattern comes for detection. The detected sequence is known as signature.

So databases should be modified or updated regularly so that new patterns can be added to database and detection can be done more effectively.

Anomaly-based detection method detects unknown attacks .as the attackers always try to choose different way to do suspicious activity so this methods works good for this type of situation .It monitors the behaviour of all the nodes in the network. If deviation of behaviour is noticed then it alerts the alarm that any malicious activity is done in network. There is no need of regular updating of database is required in this method.

There are various types of attacks which threatens the computer networks and out of which the most powerful attack is DoS (Denial of service) because it consumes all the bandwidth and resources of network and computers and temporarily deny multiple end-user services. So it is called umbrella of all the attacks. First victim was yahoo of DoS in 2000. The targets of DoS are web services and social sites.

The one which is also known as umbrella of attacks is Remote to local (R2L). Example of R2L attacks are SPY and PHF whose aim is illegal access of the network resources and they are designed to have local right permission.

Machine learning based method are much more efficient due to the better generalized property. Steps involved in ML process are:-

a) Data collection. Initially data is collected according to the proposed algorithm. For my proposed Algorithm KDD NSL dataset is taken.

b) Data pre-processing. This is the second step of machine learning process .In this step the data collected is arranged, and transformed into a format according to desired format like I have converted my dataset into CSV format so that can be easily used by the machine learning algorithm. Usually data is stored in table or array form. You doesn't required all data for your proposed algorithm so Feature extraction occurs at this step. Then Extracted data is divided into different classes for training and testing (Oscar Jimenez-del-Toro, 2017).

c) Training. After Extraction training of proposed model is done on datasets which is training datasets. 80 percent of extracted data is used for training data sets.

d) Testing. At this point training dataset is fed to model for the classification or prediction accuracy. 20 percent of datasets of extracted data used in this phase for evaluation .If the accuracy is approaches to a hundred percent

then we say that the model is better performing. At this point, comparison of results is done for best prediction of proposed algorithm.

The researcher present a framework in this paper and used some data mining algorithms i: e KNN, Decision Tree, Ensemble and hybrid algorithms to implement the intrusion detection on Host machine. Machine learning methods are utilized to improve the result performance.

The rest of paper is described as:

Section 2 described the related work which had used machine learning process in their research.

My proposed research Framework model is represented in section (3). Pre-processing of data followed for feature extraction and brief overview of training dataset and testing dataset is explained in section (4). In section (5) shows the evaluation done using training and testing datasets in research models which depicts the performance of various classifiers used in framework and attains the maximum accuracy and detection rate using hybrid classifier. Section (6) conclude the Overall research work.

2. RELATED WORK

Many studies have focused on detecting intrusion or malicious activity using machine learning algorithms. This section give some overview about the related works for implementing machine learning algorithms.

Panda et al. [1] in 2008 compared the classification algorithms, which suggest that Naïve Bayes is enchanting in terms of simplicity, robustness, elegance and effectiveness rather than other algorithms. When we consider new attacks and generalization Decision tree algorithms appeals efficient which helps construct an effective NIDS.

Adebowale et al. [2] in 2013 evaluated the classification algorithms for attack detection .They evaluated their work by using the NSL-KDD dataset. They suggested that combining more than one data mining algorithm may be used for better performance because each algorithm has its own advantages and disadvantages in terms of detection rate and different classifiers have different knowledge regarding the problem and their way to solve the problem is also different.

Nalavade et al. [3] in 2014 implemented a Network Intrusion Detection system by using Evasion technique for detecting new attacks using KDD dataset. They represented their research in which they integrate association rules for detection of intrusion. However they also proofed that by using of association rules in attack detection can create attack rules which also maintains a low false-positive rate.

Diro et al. [4] in 2017 proposed a new approach of distributed deep learning based which was based on IOT or FOG network for detect the attacks. Research shows that experimental results of new approach are better in terms of performance i: e detection rate, false alarm rate, accuracy etc. because of the sharing of parameters which can avoid local minima in training.

Othman et al. [5] in 2018 presented Spark-Chi-SVM model for intrusion detection .classification of nodes is done by using SVM classifier. Results are evaluated on Apache Spark Big Data platform with KDD99 datasets with attaining reduced training time which is efficient for big data.

Peng et al. [6] in 2018 suggested a framework model which used decision tree classifier to detect the intruder. Model was proposed over Big Data in Fog Environment. Research had implemented using KDDCUP99 dataset. They implemented model using decision tree approach with Naïve Bayesian method. The experimental results showed that the proposed method was effective and accurate.

Yavuz et al. in 2018 [7] proposed IDS system using deep-learning-based machine learning method for the detection of routing attacks for IoT. They implemented proposed research on Cooja IoT simulator for the generation attack data of high-fidelity. Experiment was done on 10 to 1000 nodes within IoT networks. They showed the good accuracy and high precision .They used hello-flood, decreased rank, and version number modification attacks for model testing.

Liu et al. in 2019 [8] had introduced a literature survey of Machine learning based and deep learning based Intrusion detection system . They also proposed a taxonomy of IDS. The

aim of survey was to clarify the concept because it fit for cyber security. Researchers explained the solution of IDS issues with that come using machine learning methods and deep learning methods.

Rupa et al. in 2020 [5] proposed A framework to detect new attacks. They used integrate data mining techniques with association rules. Process of detection was done by applying machine learning methods with CIDDS-001 dataset. Research found that applying Deep learning algorithms give better results than others with performing less computational time and less cost using

3. PROPOSED FRAMEWORK FOR INTRUSION DETECTION

This section of paper explains the overall research framework for intrusion detection by applying machine learning methods. Processing of framework done using NSL-KDD dataset. Further the KNN, Decision tree, Ensemble and Hybrid classifiers are used to detect the node whether it is attack or normal node. Framework used in this research performs excellent in terms of performance metrics.

Steps used in Research Framework are:-

I. Load NSL-KDD dataset

II. Data Extraction

III. Network Deployment

IV. Machine Learning Model Training

V. Path Selection

VI. Packet Transmission

VII. Attack Detection

VIII. Performance Analysis

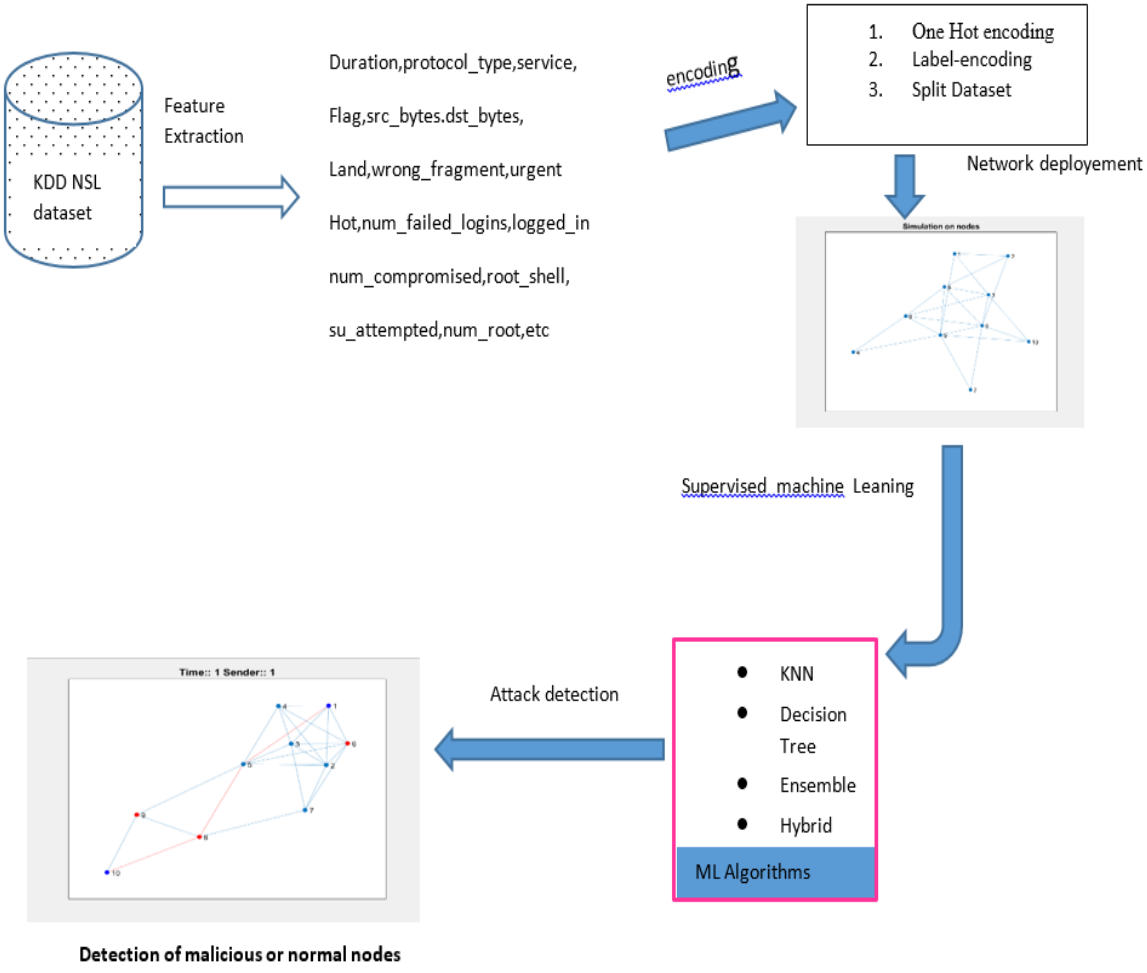


Figure 1: shows the overall Research Model Process

4. NSL-KDD DATASET PREPROCESSING AND ANALYSIS

INPUT DATASETS

To get the more useful and valuable data we always do feature extraction because dimension of data is too high for learner so we need only some valuable features only. Valuable features are always used to train the statistical model. The process where data is to be sort out in the data source according to some some rules is called as data pre-processing. To verify our model to improve the performance of system a data set known as NSL-KDD is taken.

Data pre-processing:-my original dataset contains five types of intrusions and there are 43-dimensional feature vector in each intrusion record.

Raw record of my data set is as:-

Xi= {0,tcp,ftp_data,SF,641,,0,0,,0,,0,,0,,0,,0,,1,,0,,0,,0,,0,,0,,0,,0,,0,,0,,0,,0,,0,,0,,2,,2,,0,,0,,0,,

0,,1,0,0,255,118,0.46,0.03,0.47,0,0,0.01,0,0,normal,20}

Xi is one row of dataset, there are total 43-dimensions in the original datasets and out of which 42 are the attributes of dataset and one is normal. For recording normal is kept label which records.in the pre-processing step redundancy is eliminated because many records are duplicate in the original data set which are not required. There are three more features

1. Protocol type
2. Service
3. Flag

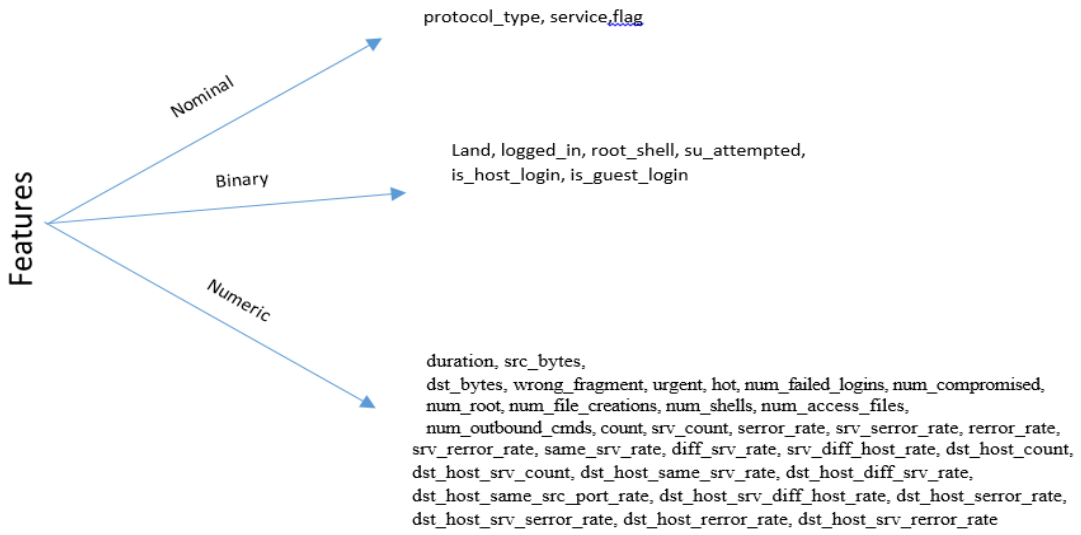


Figure 2: depicts the category wise of features of dataset

4.2. Data Pre-processing and Feature Extraction

Pre-processed of data is required to make it suitable for binary classification. For this purpose only extracted features are taken into consideration. These are some techniques used for encoding and data splitting used for data mining process.

1. One Hot encoding(one-of-k): when we have to use machine learning then data should be in the form of numerical because machine can understands only numerical data not text data. This technique is used to make all the features numerical. Transformation of all categorical features into binary features is also performed by this technique. For input the sparx matrix is used that will be of integers and attributes values should be one feature value of each attribute.

2. Label Encoding: MATLAB R2020a automatically converts label data into numerical.

3. Split Dataset: To know about the category of attacks splitting of datasets is required. Furthermore renaming is also required at this step. Renaming is as

- 0 for normal
- 1 for DOS
- 2 for Probe
- 3 for R2L
- 4 for U2R

Attacks are classified in to 5 main attacks category as the following:

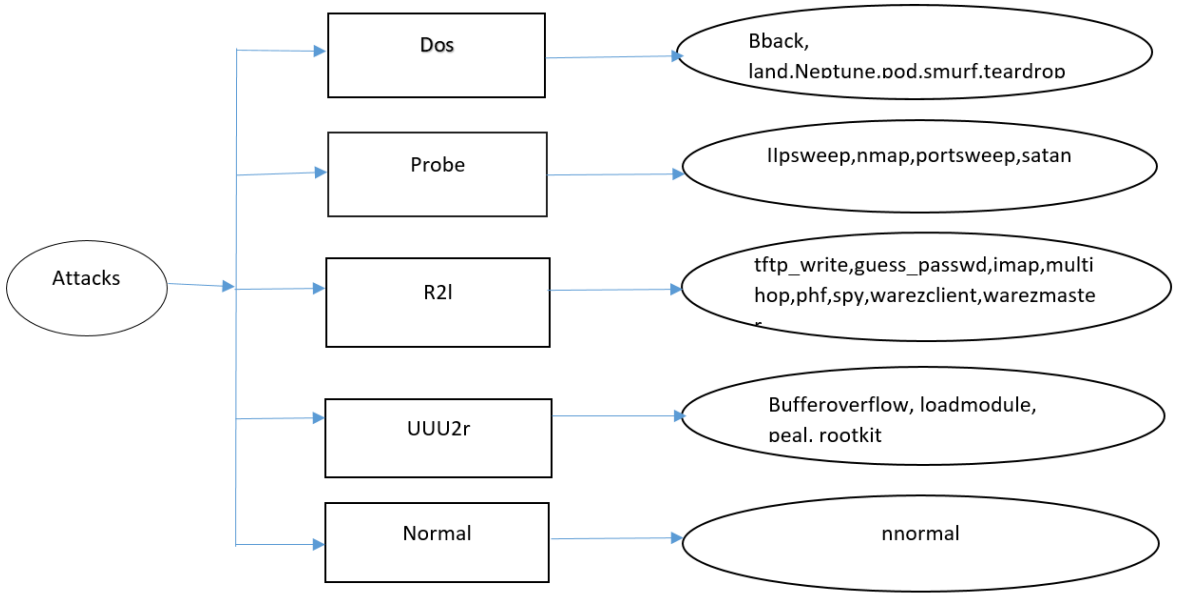


Figure 3 shows the various types of attacks and their category

4.3. Machine Learning Model Training

This phase of work use supervised learning algorithm use to train with some training datasets (80%) to give prior knowledge of malicious packets as well as normal packets.

0	tcp	ftp_data	SF	491	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	1	0	0	150	25	0.17	0.03	0.17	0	0	0	0.1	0	normal	20
0	udp	other	SF	146	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	1	0	0	0	0	0.08	0.15	0	255	1	0	0.6	0.88	0	0	0	0	0	normal	15
0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	123	6	1	1	0	0	0.05	0.07	0	255	26	0.1	0.05	0	0	1	1	0	0	neptune	19
0	tcp	http	SF	232	8153	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5	5	0	0	0	1	0	0	30	255	1	0	0.03	0.04	0	0	0	0	normal	21
0	tcp	http	SF	199	420	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	30	32	0	0	0	1	0	0.1	255	255	1	0	0	0	0	0	0	0	normal	21
0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	121	19	0	0	1	1	0.16	0.06	0	255	19	0.07	0.07	0	0	0	0	1	1	neptune	21
0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	166	9	1	1	0	0	0.05	0.06	0	255	9	0.04	0.05	0	0	1	1	0	0	neptune	21
0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	117	16	1	1	0	0	0.14	0.06	0	255	15	0.06	0.07	0	0	1	1	0	0	neptune	21
0	tcp	remote_jc	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	270	23	1	1	0	0	0.09	0.05	0	255	23	0.09	0.05	0	0	1	1	0	0	neptune	21
0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	133	8	1	1	0	0	0.06	0.06	0	255	13	0.05	0.06	0	0	1	1	0	0	neptune	21
0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	205	12	0	0	1	1	0.06	0.06	0	255	12	0.05	0.07	0	0	0	0	1	1	neptune	21
0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	199	3	1	1	0	0	0.02	0.06	0	255	13	0.05	0.07	0	0	1	1	0	0	neptune	21
0	tcp	http	SF	287	2251	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	7	0	0	0	0	1	0	0.4	8	219	1	0	0.12	0.03	0	0	0	0	normal	21
0	tcp	ftp_data	SF	334	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	1	0	0	2	20	1	0	1	0.2	0	0	0	0	warezclier	15
0	tcp	name	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	233	1	1	1	0	0	0	0.06	0	255	1	0	0.07	0	0	1	1	0	0	neptune	19
0	tcp	netbios_n	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	16	1	1	0	0	0.17	0.05	0	255	2	0.01	0.06	0	0	1	1	0	0	neptune	18
0	tcp	http	SF	300	13788	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	8	9	0	0	0	0	1	0	0.2	91	255	1	0	0.01	0.02	0	0	0	0	normal	21
0	icmp	eco_i	SF	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	16	1	0	1	1	0	0	0	0	ipsweep	18
0	tcp	http	SF	233	616	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	1	0	0	66	255	1	0	0.02	0.03	0	0	0	0	normal	21
0	tcp	http	SF	343	1178	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	9	10	0	0	0	0	1	0	0.2	157	255	1	0	0.01	0.04	0	0	0	0	normal	21
0	tcp	mtp	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	223	23	1	1	0	0	0.1	0.05	0	255	23	0.09	0.05	0	0	1	1	0	0	neptune	18
0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	280	17	1	1	0	0	0.06	0.05	0	238	17	0.07	0.06	0	0	1	1	0	0	neptune	18

Figure 4 : Sample view of Train dataset

4.4 Testing Datasets

Evaluation of model to calculate the performance is performed on 20% of datasets.

0	tcp	2	http	20	REJ	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	3	0	0	0	0	0.18	0.29	0	255	3	0.01	0.57	0.47	0	0	0	0.5	0	normal	0
0	tcp	2	domain_u	9	RSTO	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	0	0	0	0	0.11	0.56	0	255	1	0	0.27	1	0	0	0	0	0	normal	0
0	tcp	2	ecr_i	12	SF	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	0	0	0	0	0.11	0.56	0	255	1	0	0.36	0.73	0	0	0	0.2	0	normal	0
0	tcp	2	http	20	SF	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	0	0	0	0	0.11	0.56	0	255	1	0	0.24	1	0	0	0	0	0	normal	0
54	tcp	2	http	20	SF	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	4	0	0	0	0	0.29	0.21	0	255	4	0.02	0.15	1	0	0	0	0	0	normal	0
0	tcp	2	other	39	SF	10	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	4	0	0	0	0	0.21	0.21	0	255	4	0.02	0.41	0.65	0	0	0	0.3	0	normal	0
0	tcp	2	private	44	SF	10	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	2	0	0	0	0	0.1	0.43	0	255	2	0.01	0.69	0.82	0	0	0	0.2	0	normal	0
0	tcp	2	courier	3	REJ	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	40	3	0	0	0	0	0.08	0.38	0	255	3	0.01	0.58	0.99	0	0	0	0	0	normal	0
0	tcp	2	ftp	16	REJ	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	2	0	0	0	0	0.07	0.34	0	255	2	0.01	0.45	0.93	0	0	0	0.1	0	normal	0
0	icmp	1	private	44	REJ	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	3	0	0	0	0	0.23	0.46	0	255	3	0.01	0.25	1	0	0	0	0	0	normal	0
0	tcp	2	http	20	RSTR	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	82	5	0	0	0	0	0.06	0.51	0	255	5	0.02	0.47	0.87	0	0	0	0.1	0	normal	0
0	tcp	2	discard	7	SF	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	101	16	0	0	0	0	0.16	0.58	0	255	16	0.06	0.55	0.9	0	0	0	0.1	0	normal	0
0	tcp	2	domain_u	9	SF	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	11	0	0	0	0	0.11	0.61	0	255	11	0.04	0.56	0.9	0	0	0	0.1	0	normal	0
0	tcp	2	http	20	SF	10	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	2	0	0	0	0	0.18	0.45	0	255	2	0.01	0.24	1	0	0	0	0	0	normal	0
280	tcp	2	http	20	SF	10	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	4	0	0	0	0	0.29	0.29	0	255	4	0.02	0.16	1	0	0	0	0	0	normal	0
0	tcp	2	login	29	SF	10	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	3	0	0	0	0	0.27	0.45	0	255	3	0.01	0.32	1	0	0	0	0	0	normal	0
0	tcp	2	pop_3	42	SF	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	12	0	0	0	0	0.39	0.13	0	255	12	0.05	0.5	1	0	0	0	0	0	normal	0
281	tcp	2	private	44	SF	10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	134	9	0	0	0	0	0.07	0.63	0	255	9	0.04	0.56	0.89	0	0	0	0.1	0	normal	0
0	udp	3	finger	15	REJ	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32	11	0	0	0	0	0.34	0.09	0	255	11	0.04	0.58	0.99	0	0	0	0	0	normal	0
0	tcp	2	domain_u	9	RSTR	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	2	0	0	0	0	0.1	0.5	0	255	2	0.01	0.7	0.98	0	0	0	0	0	normal	0
0	tcp	2	http	20	S0	6	28	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	48	0	0	0	0	1	0	0	255	48	0.19	0.02	0.19	0	0	0	0.3	0	normal	0
0	tcp	2	domain_u	9	SF	10	28	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	83	83	0	0	0	0	1	0	0	255	83	0.33	0.02	0.33	0	0	0	0.2	0	normal	0

5. Results

5.1 performance Metrics

These are some metrics on which we have calculated performance

Detection Rate

- Detection rate or termed as "true positive rate", "recall", "sensitivity" measures the correctly detected attacks.

Detection rate $(DR) = \frac{TP}{(TP+FN)}$

False positive rate

- False positive rate also termed as false alarm rate is the true state of model and it is used as accuracy metric. Which calculates abnormal which infect are normal. Calculated as:-

False positive rate ($FPR) = \frac{FP}{FP+TN}$

Accuracy

- Accuracy is metric where the accurate fraction of results identified either attack and normal calculated as

Accuracy $(ACC) = \frac{(TP+TN)}{(TP+TN+FP+FN)}$

Precision

- Precision tells what proportion of positive predictions was actually correct.

Precision is also called as positive predictive value tells us what proportion of positive attacks are actually indeed attacks. Precision is Appropriate when are focusing on minimizing false positives.

Precision = $\frac{TP}{TP+FP}$

Fmeasure

- Fmeasure is the harmonic mean of precision and recall value with giving equal weightage on both.

Fmeasure=$\frac{(2*precision*recall)}{(precision+recall)}$

- Confusion Matrix

There will be a Confusion matrix which is useful for measuring the performance of our model to solve intrusion detection classification problem.

- True positive (TP): TP can be calculated as no. of exact instances which are classified as an intrusion.

• True Negative (TN): TN gives inaccurate no. of instances termed as intrusion.

• False-positive (FP): suggest in no. instances of intrusion instances that were wrongly considered as normal.

• False-negative (FN): gives no. of normal instances which are mistakenly classified as an attacks.

Table 1 shows the confusion matrix

Activity	Predicated	Normal
Attack	True positive(TP)	False Negative(FN)
Normal	False Positive(FP)	True Negative(TN)

5.2 Results and discussion

5.2.1 Applying Classification Methods

This section demonstrates that model is tested on varying of no. of nodes present in network. Result is calculated in three scenarios i: e 10, 20 and 50 no. of nodes. Thereafter various classification techniques are applied in all cases by changing no. of nodes. Classifier algorithm such as Decision tree, KNN, Ensemble and Hybrid algorithms are applied to check the attacker class or normal class. Fig 6 shows the detection of suspicious node after applying these classifier algorithms. Fig 7 depicts the confusion matrix in which TP, FN, FP, and TN are shown and different diagrams in fig 7 also reveals the comparison by applying different classifiers. Furthermore it also compares the recall value, precision value, accuracy, fmeasure, detection rate, false alarm rate in table 2 when testing is done by changing techniques for classifying nodes.

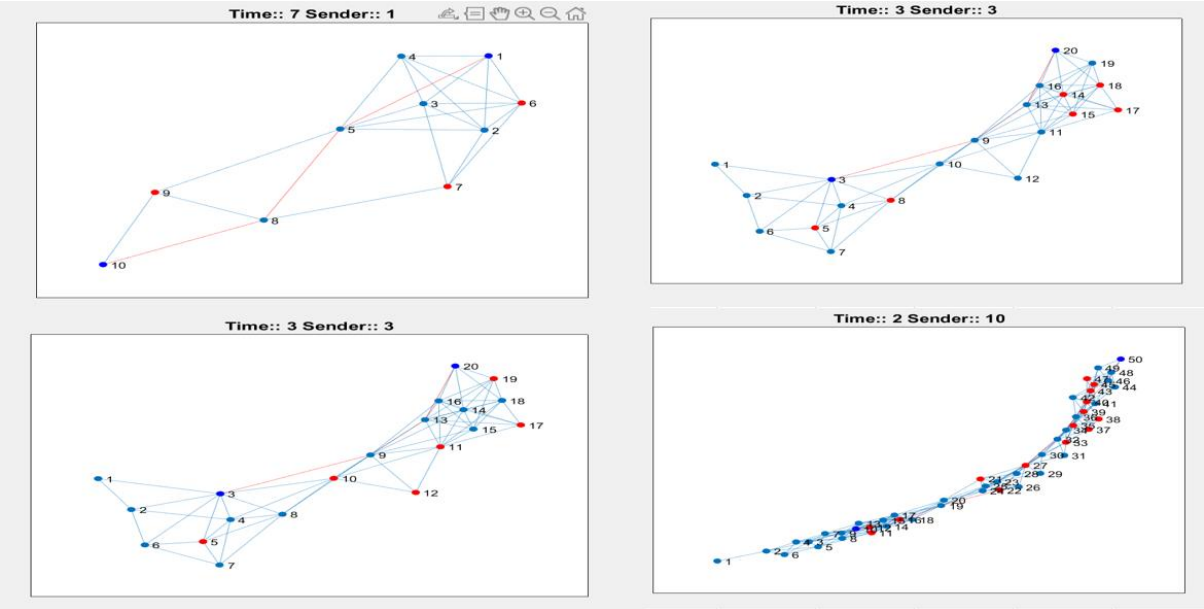


Figure 6 shows that 10,20,50 nodes are communicating at time .fig shows that blue nodes are normal nodes where red colored nodes may be malicious .Model used in this research is trained on 80% of data so on the basis of training the host machine will detect the malicious nodes.Everytime when you run research model then every time a new network is made by Nodes so sender nodes are different at different time i:e sender node is 1 when time is 7 and sender node is 3 when time is 3 and sender is 3 when time is 3 but at this time suspicious nodes are changed from 8 to 10 and 12 and sender is 10 when time is 2.further research has applied shortest distance path algorithm to detect the path between the source to destination if any malicious node comes under the selected path from source to destination at that time the that malicious node will be detected by host machine.

Every time when we run our research model then every time a new path will be chosen.

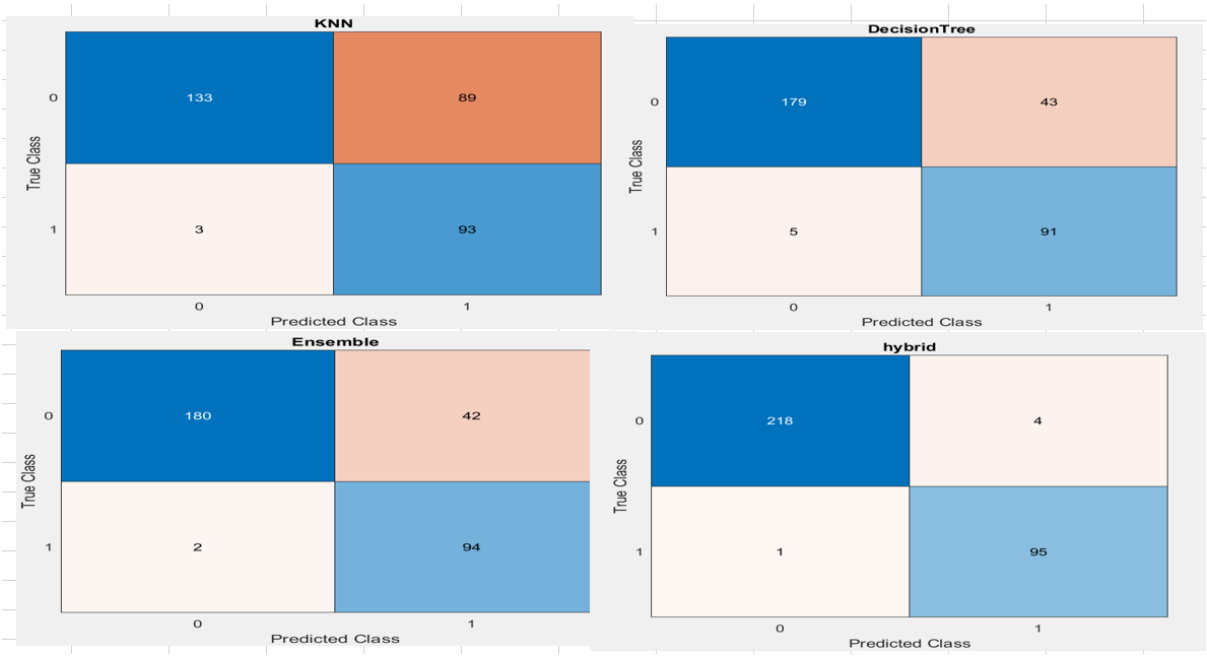


Figure 7 shows the confusion matrix of KNN, Decision Tree, and ensemble and Hybrid classifiers

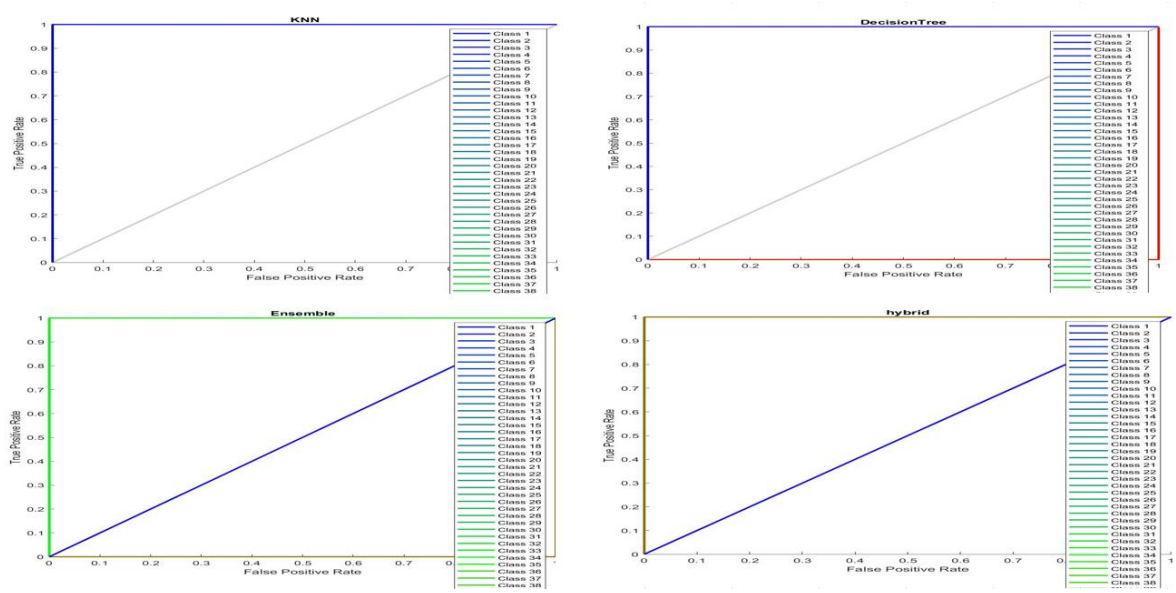
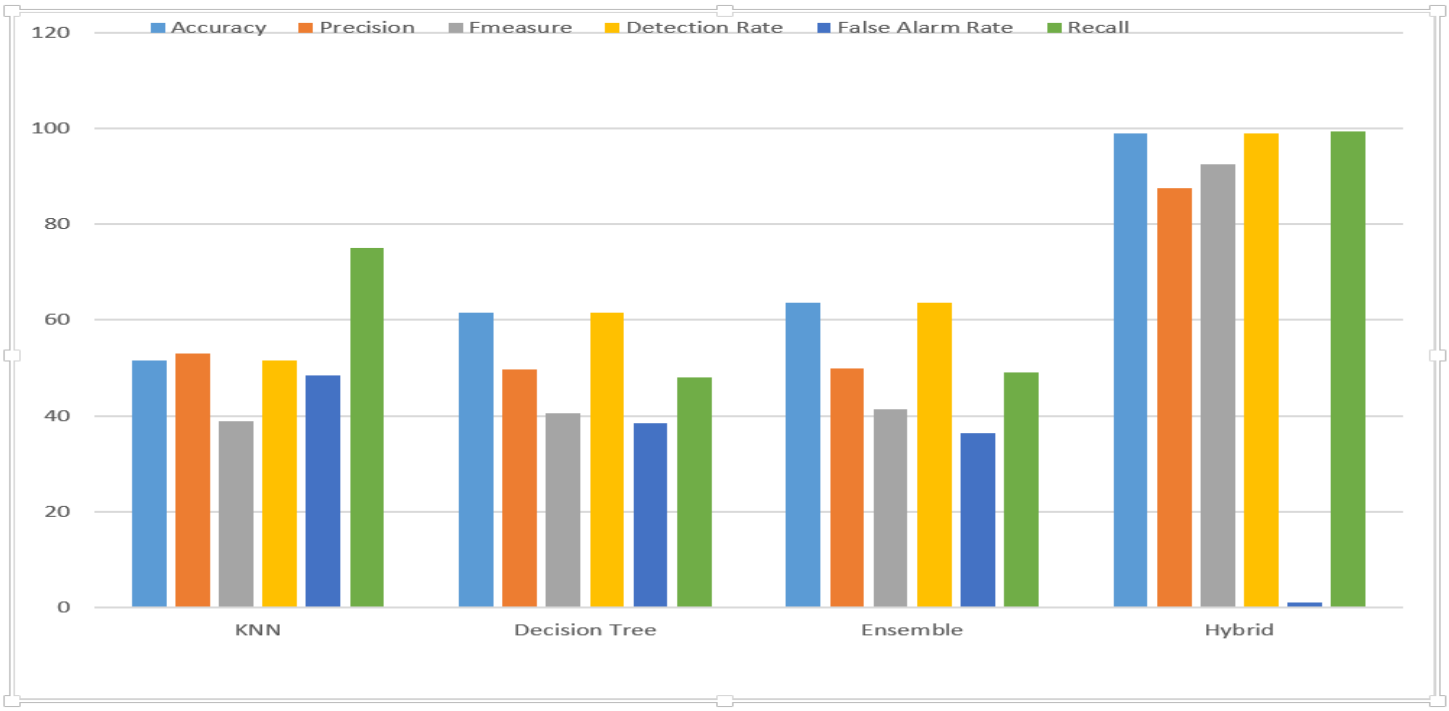


Figure 8 shows the ROC curve of true positive rate of KNN, Decision tree, Ensemble and Hybrid Classifiers

Table 2 depicts the performance of all classifiers based on Accuracy,Precision,recall,Fmeasure,Detection Rate,False Alarm Rate and according to the table hybrid classifier has highest accuracy of 98.99 and false alarm rate is very low to others as it 1.01. According to the table Hybrid classifier outperforms the others in all other metrics also

Classifier name	Accuracy	Precision	Recall	Fmeasure	Detection Rate	False Alarm Rate
KNN	51.52	52.94	75.00	38.89	51.52	48.48
Decision Tree	61.62	49.74	47.92	40.47	61.62	38.38

Ensemble	63.64	49.87	48.96	41.38	63.64	36.36
Hybrid	98.99	87.50	99.48	92.60	98.99	1.01



Graph 1: illustrate the performance of all classifier on the basis of performance metrics

6. CONCLUSION

Research has been going on in Network Intrusion Detection for the privacy, security and integrity of data. The aim of all researcher in the same is to minimize false positive rate and always try to maximize the accuracy rate. This research introduces a framework model which uses hybrid classifier which is the integration of data mining classification techniques to implement Host intrusion detection using machine learning process. Efforts are applied using hybrid classifier that achieves goal by giving best results when evaluation on i: e accuracy, false alarm rate, detection rate, Recall, fmeasure, and Precision values are taken. This research model is tested by considering unknown attacks. The research framework is implemented using NSL-KDD datasets. Three scenarios of 10, 20 and 50 nodes are taken for performance testing.

References

[1] A M. Panda and M. R. Patra, (2008), A comparative study of data mining algorithms for network intrusion detection, in First Int. Conf. on Emerging Trends in Engineering and Technology, IEEE,Nagpur, Maharashtra.

[2] A. Adebowale, S. A. Idowu and A. Amarachi(2013), Comparative study of selected data mining algorithms used for intrusion detection, International Journal of Soft Computing and Engineering, vol. 3, no. 3, pp. 237–241.

[3] K. Nalavade and B. B. Meshram (2014), Mining association rules to evade network intrusion in network audit data, International Journal of Advanced Computer Research, vol. 4, no. 2, and pp. 560–567.

[4] A. A. Diro and N. Chilamkurti(2017), distributed attack detection scheme using deep learning approach for

Internet of Things, Future Generation Computer Systems, vol. 82, pp. 761–768,.
[5] S. M. Othman, F. M. Ba-Alwi, N. T. Alsohybe and Y. A. Amal(2018), Intrusion detection model using machine learning algorithm on big data environment, Journal of Big Data, vol. 5, no. 1, pp. 521.
[6] K. Peng, V. C. M. Leung, L. Zheng, S. Wang, C. Huang et al.(2018), Intrusion detection system based on decision tree over big data in fog environment, Wireless Communications and Mobile Computing, vol. 2018, no. 5, pp. 1–10.
[7] F. Y. Yavuz, D. Ünal and E. Gul(2018), Deep learning for detection of routing attacks in the Internet of Things, International Journal of Computational Intelligence Systems, vol. 12, no. 1, pp. 39–58.
[8] H. Liu. and B. Lang(2019), Machine learning and deep learning methods for intrusion detection systems: A survey, Applied Sciences, vol. 9, no. 20, pp. 4396.
[9] T. Rupa Devi and S. Badugu(2020), A review on network intrusion detection system using machine learning, in Int. Conf. on Emerging Trends in Engineering 2019, LAIS, Switzerland: Springer Nature Switzerland, vol. 4, pp. 598–607.
[10] A. Sahasrabuddhe, S. Naiade, A. Ramaswamy, B. Sadliwala and P. R. Futane(2017), Survey on intrusion detection system using data mining techniques, International Research Journal of Engineering and Technology, vol. 4, no. 5, pp. 1780–1784.
[11] L. Dali, A. Bentajer, E. Abdelmajid, K. Abouelmehdi, H. Elsayed et al.(2015), A survey of intrusion detection system, in 2nd World Sym. on Web Applications and Networking, Tunisia, Piscataway: IEEE, pp. 1–6.
[12] V. Veeralakshmi and D. Ramyachitra(2015), Ripple down rule learner (RIDOR) classifier for iris dataset, International Journal of Computer Science Engineering, vol. 4, no. 3, pp. 79–85.
[13] S. Brugger(2011), Data mining methods for network intrusion detection, Ph.D. dissertation, University of California, Davis, pp. 1–65, 2011.
[14] Vipin Kumar, Himadri Chauhan, Dheeraj Panwar(2013), K-Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume- 3, Issue-4, September .
[15] L. Dali, A. Bentajer, E. Abdelmajid, K. Abouelmehdi, H. Elsayed et al.(2015), A survey of intrusion Detection system," in 2nd World Sym. on Web Applications and Networking, Tunisia, Piscataway: IEEE, pp. 1–6.