International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

Machine Learning Based Classification Model for Network Traffic Anomaly Detection

Dr. K. Shyam Sunder Reddy[1], Dr. Vempati Krishna[2], Dr. M. Prabhakar[3], Punna Srilatha[4], Dr. K.Gurnadha Gupta[5], Ravula Arun Kumar[6]

[1]Assoc.Professor, Dept of IT, Vasavi College of Engineering (A), Hyderabad
[2]Professor, Dept of CSE, TKR College of Engineering and Technology (A), Hyderabad
[3]Assoc.Professor, Dept of IT, Vignan Institute of Technology and Science, Hyderabad
[4]Asst.Professor, Dept of IT, Vignan Institute of Technology and Science, Hyderabad
[5]Assistant Professor,
Department of CSE, Koneru Lakshmaiah Education Foundation,K L Deemed To Be University, Vaddeswaram, AP, India.
[6]Assistant Professor, Department of CSE, Vardhaman College of Engineering(A), Hyderabad

Abstract— In current days, cloud environments are facing a huge challenge from the attackers in terms of various attacks thrown to the cloud service providers. In both industry and academics, the problem of detection and mitigation of DDoS attacks is now a challenging issue. Detecting Distributed Denial of Service (DDos) threats is mainly a classification problem that can be addressed using data mining, machine learning and deep learning techniques. DDoS attacks can occur in any of the seven-layer OSI model's network. Hence, detecting the DDoS attacks is an important task for cloud service providers to overcome dangerous attacks and loss incurred to stake holders and also the provider..

Keywords- DDos, OSI, machine learning**.**

I. INTRODUCTION

Denial-of-service (DoS) and disbursed denial-of-service (DDoS) assaults have turn out to be an increasing number of famous in current years, with attackers sending a huge wide variety of packets to the valid person device as a way to make on line structures unavailable for them. DoS assaults are venomous operations that save you legal customers from getting access to a device, community, software, or records. The modern instances have simply validated the growth in cyberattacks. There are mostly styles of disbursed dos assaults: unmarried supply assaults, which originate from a unmarried device, and disbursed dos assaults, which originate from more than one structures. DDoS (Distributed Denial of Service) assaults are blanketed withinside the DoS category. DDos assaults pick to goal Internet infrastructure, routers, DNS servers, bandwidth, and servers, amongst different things .DDoS volumetric assaults account for greater than 65% of all assaults of this kind.

The attackers regularly use those strategies, together with social media, email, and internet apps, to perform the assault and transmit the code that infects the device. They every now and then appoint the approach of making and making use of a botnet, additionally called a community of hijacked machines. If their assault is successful, the attacker could be capable of manage the gadget in any manner they want. All of this could be completed from wherein they may be sitting with out the device's proprietor even knowing.

Anomaly detection is a first-rate subject for community customers in trendy technological era. Users of the community also are growing, this means that there's greater visitors at the community because of the improvement of numerous community strategies. This makes it very difficult to identify uncommon styles. The framework/model's accuracy degree changed into additionally mentioned on this paper, at the side of a top level view of the numerous ML strategies used to resolve the ambiguity detection trouble and their benefits and disadvantages. Strategies for figuring out and mitigating community visitors anomalies are mentioned and as compared on this survey in phrases of accuracy and kind of anomaly. Network visitors anomaly detection studies gaps and essential questions are mentioned in detail. We expect that the researchers could be guided withinside the proper path for project superior studies on this vicinity through the evaluation, comparisons, and next identity of gaps.

Digital transformation, digitization, enterprise 4.0, etc.are the buzzwords, and the primary purpose is to apply era and records to enhance accuracy, productivity, and efficiency. The Key enabler is capable of extract beneficial records from a huge quantity of records, allowing capability optimization, time savings, and price reduction. There are plenty of strategies used to extract beneficial records. Data is analyzed the use of plenty of strategies as a part of the records evaluation process. The fine of offerings and safety in huge-scale networks have offered a regular project to numerous community groups in current instances. These safety problems may be delivered on through plenty of inner or outside elements. External elements encompass stealing safety records or shutting down all offerings; inner elements encompass configuration errors,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

visitors congestion, energy outages, and server crashes. In addition to all of those problems, a unmarried safety threat, additionally called an anomaly, is presently widespread. The styles of numerous datasets fluctuate from the ones of the everyday dataset, or the records deviated from the ordinary dataset at instances. Anomaly is the call given to this deviation, which has a bad effect on community operations and has a large effect on community offerings. There are many approaches to outline anomaly.Lakhina et al. nation that [1], "anomalies are distinct styles and a moderate extrade withinside the visitors ranges of a community".

II. LITERATURE REVIEW

Based on gadget getting to know and deep getting to know techniques, the subsequent are the principle thoughts for community IDPS work. In current decades, speedy improvements in generation and networks have resulted from the tremendous usage of Internet offerings throughout all industries. Since counterfeiting has elevated and several current structures were compromised, it's miles now crucial to increase statistics protection answers that may pick out new assaults.

Et., M. Mithem al detected unknown assault programs the use of a deep neural community the use of an progressive intrusion detection gadget with extraordinary community performance [4]. Attacks may be detected the use of both binary class or multiclass class. The recommended methods produced promising results in phrases of excessive accuracy. Security is a massive problem nowadays due to the fact each hour, a whole lot of statistics is exchanged throughout all domains .Data protection may be safeguarded the use of a neural community withinside the way defined in [5].Various IDS and Intrusion Prevention Systems (IPS) have been tested and in comparison on this have a look at.In addition, a contrast of numerous methods is carried out. Various fashions and techniques for intrusion detection also are mentioned here. The have a look at suggests how associated troubles may be solved with neural networks.

Chewet et al.2020) This paper's tree policies NIDS have been capping a position to triumph over their visibility problems in [10]. A selection tree-primarily based totally amendment is made to the pruning algorithms. By deciding on the maximum essential policies, this framework preserves privacy, and any small adjustments haven't any impact at the method choice approach however do have an effect on the gadget's performance. This paper with the aid of using Bhatiet al. (2020) makes a framework that is largely applied with the assist of the MATLAB software. An character classifier is evolved and skilled inside this framework, and it makes a decisive selection primarily based totally on majority vote. The selection is the remaining of the 4 most important steps on this framework, which encompass statistics collection, pre-processing, training, and testing. On numerous datasets, it gives excessive detection accuracy. The paper's disadvantage is that it gives a complex framework structure.

D'Souza, D. J., and others2021) Outlier detection in unstructured statistics changed into the concern of this paper [12].t may be represented the use of a graph. A survey of static, dynamic, and gadget getting to know methods to anomaly detection is likewise protected on this paper, with a more emphasis on graph-primarily based totally techniques. The gadget will become greater complex due to its use of more than one graph scenarios, that is a disadvantage.

The aim of the work [8] is to apply deep getting to know-primarily based totally intrusion detection and prevention strategies that may at once prevent assaults like DOS, R2L, and U2R.The intrusion changed into visualized the use of an in-intensity getting to know version, that is a multi-degree comprehension skilled with excessive precision withinside the kddcup99 dataset. In order to estimate the bring about actual time, the Display Deep Learning version collects the perfect community statistics and saves it as a CSV file. In the second one step, the intrusion is avoided with the aid of using heritage scripts.The script's goal is to finish the prevention section with the aid of using recommending numerous assault-unique preventative measures. The Multi-Layer Perceptron class component's statistics may be used to make a selection. Specialized intrusion detection and prevention structures are blended right into a unmarried gadget to facilitate quicker and greater green intrusion detection and prevention.

Following an evidence of IDS, the have a look at [10] provides a class primarily based totally at the maximum not unusual place methods to the advent of Network-primarily based totally IDS (NIDS) structures, inclusive of gadget getting to know and deep getting to know. The present day NIDS-primarily based totally research are reviewed in intensity, with an emphasis on proposed answers, benefits, and drawbacks. The proposed approach, assessment criteria, and dataset choice are all mentioned after current ML and DL-primarily based totally NIDS improvements and traits are mentioned. The flaws in the approaches that were presented were used to highlight a number of research issues as well as the potential for further investigation into enhancing ML- and DL-based NIDS. Some of the advantages and disadvantages of various methods, as well as difficulties in building a model.

Challenges Identified:

- Examining the characteristics of attacks—such as their low or high rate—as well as the security issues brought on by the diversity of connected objects
- Identifying the attack as a specific kind of attack.
- Developing techniques for spotting attacks

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

- Finding the balance between academic propositions and the industrial practice of combating DDoS.
- Overcoming the loss of money.
- Authentication, user privacy, and data leakage remain major obstacles for cloud computing environments.

III. METHODOLOGY

The below are the research methodologies previously used

- **FFSC SCORES METHOD**

Step-1: Load the dataset
Step-2: Drop the rows with null,infinite values
Step-3: Normalize the dataset and multiply with 255
Step-4: Convert the dataset into integer type
Step-5: Find the mean of each row
Step-6: Feature Feature ordered Relation (FFoR): FFoR of a feature fi with all other features fj of an object Oi is defined using below Equation

$$FFoR\left(O_i^{f_i}\right) = \sum_{j=1\&i\neq j}^{n} \left(|f_i - f_j|\right)$$
$$where\ l \leq i \leq n.$$

Step-7: Average FFoR (AFFoR): We define AFFoR of an object Oi as the mean value of its individual FFoR values and it can be expressed using below Equation

$$AFFoR(O_i) = \frac{\sum_{j=1}^{n} (\text{FFoR}(O_i^{f_j}))}{n}$$

Step-8: The Deviation vector (Dev): Deviation vector of an object O i can be defined as the absolute difference between the FFoR values of the object and its corresponding AFFoR value. The Dev of a feature fj is computed using below Equation

$$Dev\left(O_i^{f_j}\right) = lAFFoR(O_i) - FFoR(O_i^{f_j})|, \forall j = 1,2,\dots n$$

Step-9: FF-score (FFSc): We define FFSc of an object Oi as the degree of similarity in terms of its Dev and mean value, which is given by the below Equation.

$$FFSc(O_i) = \frac{(O_i \times Dev(O_i)T)}{(mean(O_i) + mean(Dev(O_i)))}$$

- **Clustering with E Power Distance Method**

Step-1: Load the dataset
Step-2: Drop the rows with null, infinite values
Step-3: Normalize the dataset by dividing each value with maximum value of that respective class label column.
Step-4: Find Mean and Standard Deviation of each column in the dataset.
Step-5: Compare mean values with values in the dataset.There are mainly three condition in numerator:

1.If both the values are not equal to 0 then 0.5(1+e-(meani-datai/σi)2)
2.If any of the value is 0 then -1
3.Else 0

Step-6: Compare mean values with values in the dataset.There are mainly three condition in denominator:

1.If both values are 0 then 0
2.Else 1

Step-7: Favg=ΣNumerator/ΣDenominator
Step-8: Sim=1+Favg/2

The proposed method presented here overcomes all of the issues and restrictions raised. On the intrusion set dataset, the experimental results of various machine learning classification algorithms are examined in this section. Unsupervised and supervised learning are the two most common approaches to machine learning. For training algorithms, labelled examples such as an input with a chosen output are used. Instances without labels are trained through unsupervised learning. Exploring the data and finding some structure in the data are the two objectives of unsupervised learning. Semi-supervised learning and reinforcement learning are also used [12] in addition to these approaches.

A collection of records is referred to as a data set. The csv files are the kind of data we're using here. The abbreviation for "comma separated value" is "csv." We need a data set in order to acquire the model and train it. We begin with the csv data in this paper. It will be transformed into images in the future, and those images serve as the model's input. We have utilized the CIC Ddos 2017, CICDdos 2019, and NSL-KDD data sets in this case.

Recognizing, comprehending, and categorizing objects into predetermined "sub-populations" is the process of classifying them. The ML algorithms classify future datasets using a variety of algorithms and pre-categorized training datasets. In machine learning, classification algorithms use training data to predict whether new data will be classified as descriptive. To put it succinctly, classification is a subset of "pattern recognition" that employs classification techniques in training data to locate a pattern in the data sets [13].The following classification algorithms were used to identify and classify intrusive attacks in this case: Random Forest, AdaBoost, Extra Trees, Gradient Boost, Linear Regression, and Multilayer Perceptron.

First and foremost, we must understand and visualize each feature; however, it is extremely challenging to analyze and visualize all of these features. Therefore, the first thing we need to do is reduce the number of features in the data set and ensure that only the most crucial ones are included. Verify that none of the features are duplicates or outliers. We must apply the Andrew curves concept to the data set after removing the undesirable features in order to clearly visualize the data set. We can convert high-dimensional data to two-dimensional data using Andrew curves, which is very convenient.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

When working with a machine learning model, the part of the feature selection process where we make sure we have the right features to train the model well is very important. High-dimensional features, on the other hand, present a challenge when training a large number of features because it is difficult to visualize high-dimensional data. Andrew curves are utilized here.

The Andrew curve is used to display large amounts of data. To process them, we need to properly visualize the high-dimensional features in our dataset. This is why we need the Andrew curve. The Andrew curve helps determine whether the data are highly linear, nonlinear, or linear.

Andrew curves are need for the following reasons:

- To visualize high dimensional data.
- To understand the behaviour of the data set.
- To analyse the approach and the methodology.
- To make sure if we have to use the supervised or unsupervised algorithms.

We can only visualize the data which is 2 dimensional but when we have the data set with more dimensions we need the Andrew curves to visualize that data set dimensions. Andrew curves are used to envision high dimensional data. And by mapping each observation onto a function, in that all Means, distances, and variances were preserved. It is calculated using the following formula:

T(n) = x_1/sqrt(2) + x_2 sin(n) + x_3 cos(n) + x_4 sin(2n) + x_5 cos(2n) + …

The plotting module's Andrews curves () method can be used to plot Andrews curves on a graph.Multivariate data clusters can be visualized by using the matplotlib plot of Andrews curves that is generated by this program.

The following are some of the attack categories identified in the datasets:

- DoS (Denial of Service): DoS is a type of intrusion attack in which network resources are overloaded and made unavailable to authorized users.
- Root to User (U2R): By granting the invader root access, this is an intrusion attack that puts the user's credibility in jeopardy.
- Local to Remote (R2L): It is an attempt to break into a network when the integrity of the network is compromised, giving the attacker access to the local network.
- Probe: An intrusion activity known as a probe entails scanning the network and gathering any and all information pertaining to network activities.

The various Network Anomalies and Network Attacks implemented in this paper are depicted in Fig. 1 below.

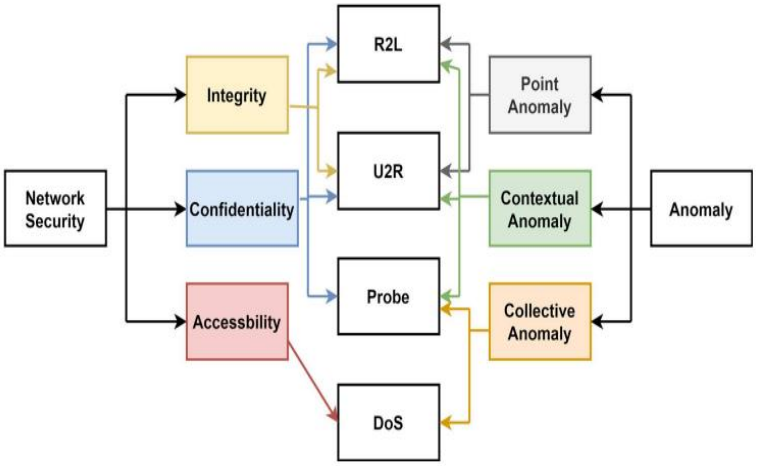


Fig. 1. Network Anomalies and Network Attacks

IV. IMPLEMENTATION

Due to the highly nonlinear nature of the datasets, a variety of Machine Learning and Deep Learning methods can be used to quickly identify normal traffic from malicious traffic. A new method for correctly distinguishing traffic is developed by combining various deep learning and machine learning approaches. This system can be used in cloud, FOG, SDN, and IoT environments to detect attacks and safeguard the environment.

In this paper, along with other datasets mentioned, primarily we implemented the dataset Functional Description is CIC-DDoS2019. The description of the dataset is given below.

We have selected the CICDdos 2019 data set to carry out this task. When we followed the link to http://205.174.165.80/CICDataset/CICDDoS2019/, we discovered that there are two distinct file types: These are known as Pcaps, or packet capture files, and they contain the raw data from a two-day experiment. CSVs, or comma-separated value files, are a type of file in which the data is stored in csv format and the type of attack is manually labeled. In the beginning, we directly considered using cvs as an input to the model. When the data in the csv files is complete and clean, we can only use them as input for the model. However, there are a few issues with it: The few records that are missing and the infinity of records are inapplicable to deep learning algorithms. Another issue is that only a few of the data's columns contain any significant figures. As a result, we are once more regenerating the csv files from the pcap files in order to address this issue.which CICFlowMeter enables us to carry out.

CICFlowMeter, developed by the Canadian Institute for CyberSecurity, can generate approximately 83 features from Packet Capture Files and review those features.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

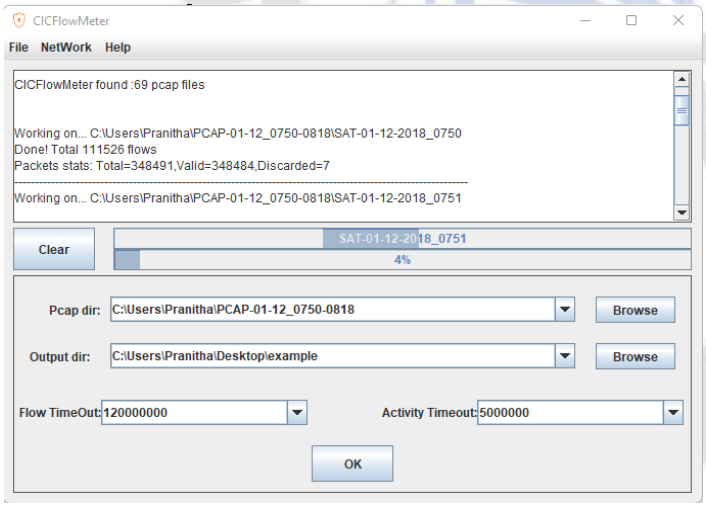
- Step-1: csv files to pcap (packet capture) files.

We need to convert pcapto csv files in order to carry out feature analysis, data cleaning, and analysis.At first, the pcap files contained information such as time, source, destination, protocol, length, and so on.Now, this pcap file needs to be converted into a csv file with approximately 80 attributes and daily values.

- Data Set Creation from Pcap File To Image Equivalent: Procedure to convert pcap to csv files
- Run command prompt as '**adminimumistrator**'. -press **windows+R** to open "Run" box. Type "cmd" into the box and then press **ctrl+shift+enter**. By running the below commands we can convert pcap to csv files.
- -C:/Windows/System32>cd.. (to move out of the existing directory). -C:/Windows>cd… -C:/cd CICFlowMeter. -C:/CIC FlowMeter> cd bin. -C:/CIC FlowMeter\bin >CICFlowMeter.bat (This opens CICFlowMeter application window).

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	172.16.0.5	192.168.50.1	IPv4	183	Fragmented IP protocol (prot
2	0.000002	172.16.0.5	192.168.50.1	UDP	1514	588 → 3191 Len=1621
3	0.000003	172.16.0.5	192.168.50.1	IPv4	183	Fragmented IP protocol (prot
4	0.000155	192.168.50.7	4.2.2.4	DNS	84	Standard query 0xbaa2 AAAA a
5	0.000157	172.16.0.5	192.168.50.1	IPv4	1514	Fragmented IP protocol (prot
6	0.000158	172.16.0.5	192.168.50.1	UDP	129	589 → 19033 Len=1567
7	0.000159	192.168.50.7	4.2.2.4	DNS	84	Standard query 0xbaa2 AAAA a
8	0.000160	172.16.0.5	192.168.50.1	IPv4	1514	Fragmented IP protocol (prot
9	0.000161	172.16.0.5	192.168.50.1	UDP	129	589 → 19033 Len=1567
10	0.000239	172.16.0.5	192.168.50.1	IPv4	1514	Fragmented IP protocol (prot

Fig. 2. Sample CSV file of Pcap



	A	B	C	D	E	F	G	H
1	Unnamed: (	Flow ID	Source IP	Source Por	Destinatior	Destinatior	Protocol	Timestamp F
2	425	172.16.0.5-	172.16.0.5	634	192.168.50.	60495	17	51:39.8
3	430	172.16.0.5-	192.168.50.	634	172.16.0.5	60495	17	51:39.8
4	1654	172.16.0.5-	172.16.0.5	634	192.168.50.	46391	17	51:39.9
5	2927	172.16.0.5-	172.16.0.5	634	192.168.50.	11894	17	51:39.9
6	694	172.16.0.5-	172.16.0.5	634	192.168.50.	27878	17	51:39.9
7	838	172.16.0.5-	172.16.0.5	634	192.168.50.	47149	17	51:39.9
8	3090	172.16.0.5-	172.16.0.5	634	192.168.50.	22713	17	51:40.0
9	2594	172.16.0.5-	172.16.0.5	634	192.168.50.	49912	17	51:40.0
10	2698	172.16.0.5-	172.16.0.5	634	192.168.50.	56681	17	51:40.0
11	277	172.16.0.5-	172.16.0.5	634	192.168.50.	13161	17	51:40.0
12	745	172.16.0.5-	172.16.0.5	634	192.168.50.	25051	17	51:40.0
13	1381	172.16.0.5-	172.16.0.5	634	192.168.50.	12606	17	51:40.1

Fig. 3. Sample CSV file of Pcap

Upon running the above command Fig 3, when a pcap file is taken as input each packet from the pcap file is loaded and equivalent column separated values.In the end csv file will be generated. The generated csv file in Fig 3 has attributes like unnamed,Flow ID,Source IP,Destination Port,Destination IP,Source Port,Protocol,Flow Duration etc.

Normalization for one channel images: By applying the below normalization formula to the above instances, we get instances data, Later Converting them into Integer values, and generate Instances and Images

$$X' = \frac{X - Min(X)}{Max(X) - Min(X)} \times 225$$

	Flow Dura	Total Fwd	Total Leng	Fwd Packe	Fwd Packe	Fwd Packe	Flow Pack	Flow IAT I	Fwd IAT Tc	Label
2	1	2	766	383	383	383	2000000	1	1	UDP-lag
3	1	2	778	389	389	389	2000000	1	1	UDP-lag
4	2	2	750	375	375	375	1000000	2	2	UDP-lag
5	40694755	5	1500	300	300	300	0.122866	10173689	40694755	Benign
6	1.14E+08	52	0	0	0	0	0.456655	2232776	1.14E+08	Benign
7	1.13E+08	39	0	0	0	0	0.345327	2972008	1.13E+08	Benign
8	1.19E+08	60959	0	0	0	0	515.4884	1939.939	1.19E+08	MSSQL
9	1	2	2944	1472	1472	1472	2000000	1	1	MSSQL
10	2	2	2944	1472	1472	1472	1000000	2	2	DrDoS_SSDP

Table 1: Generation of Instances and Images

Instance 1	0	0	66	
	66	66	66	
	255	0	0	
Instance 2	0	0	67	
	67	67	67	
	255	0	0	
Instance 3	0	0	64	
	64	64	64	
	127	0	0	
Instance 4	87	0	129	
	51	51	51	
	0	255	87	
Instance 5	243	0	0	
	0	0	0	
	0	55	243	
Instance 6	241	0	0	
	0	0	0	
	0	74	241	
Instance 7	255	255	0	
	0	0	0	
	0	0	255	
Instance 8	0	0	255	
	255	255	255	
	255	0	0	
Instance 9	0	0	255	
	255	255	255	
	127	0	0	

Syntax: andrews_curves(frame, class_column, ax=None, samples=200, color=None, colormap=None,**kwargs)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

Understanding the Nature of Datasets Using Andrew Curves: Andrew Curves for DDoS Evaluation Dataset (CIC-DDoS2019)

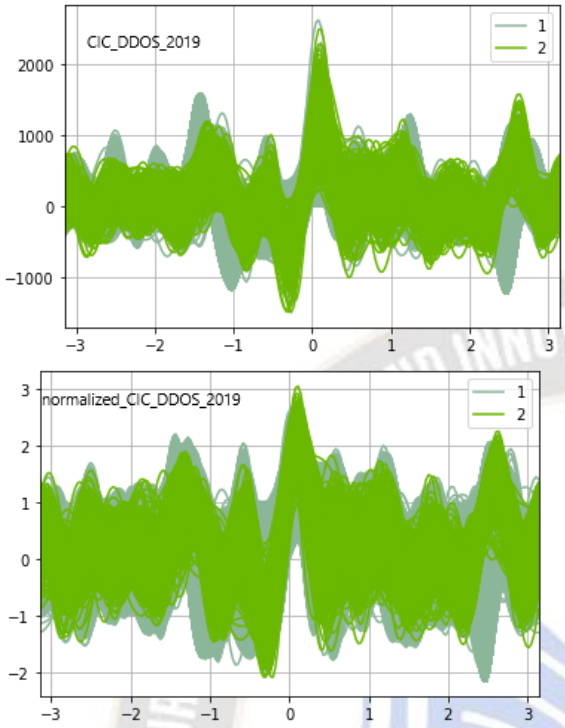


Fig.4. Andrew curve for CIC DDoS 2019 and Normalized CIC DDoS 2019

Andrew Curves For NSL-KDD Dataset: DoS, Normal, Probbing, Remote to local(R2L), User to Root(U2R).

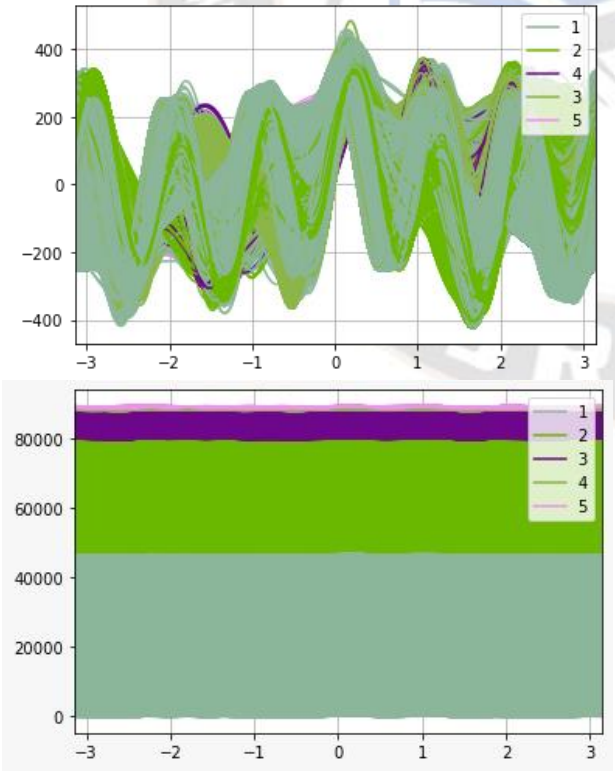


Fig. 5. Andrew curve for NSL-19 and Normalized NSL-19

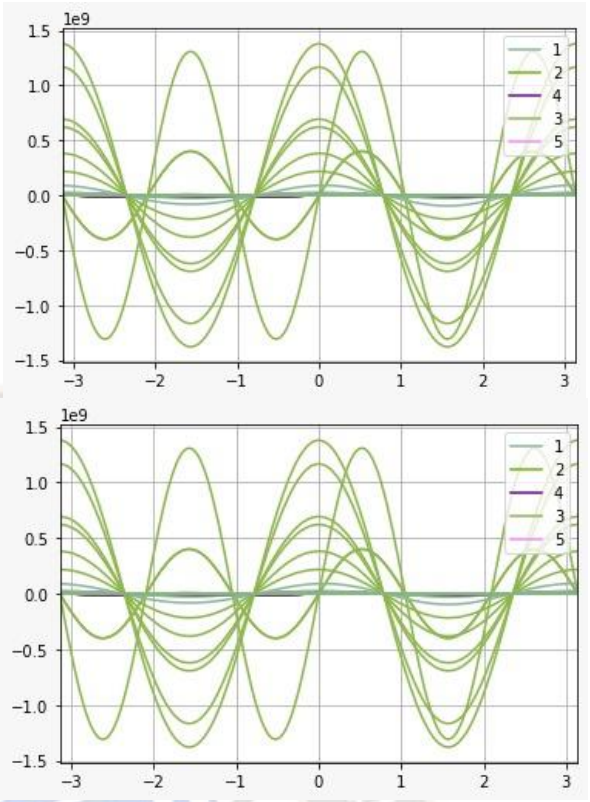


Fig. 6. Andrew curve for NSL-41 and Normalized NSL-41

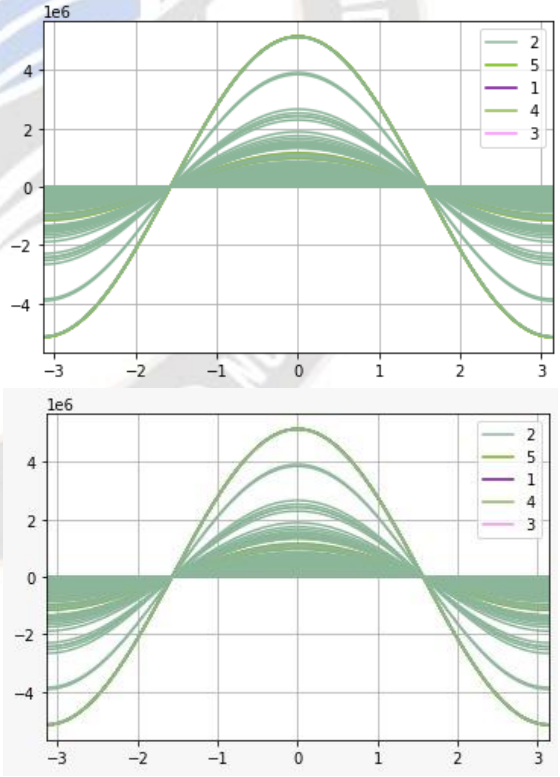


Fig. 7. Andrew curve for KDD and Normalized KDD-19

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

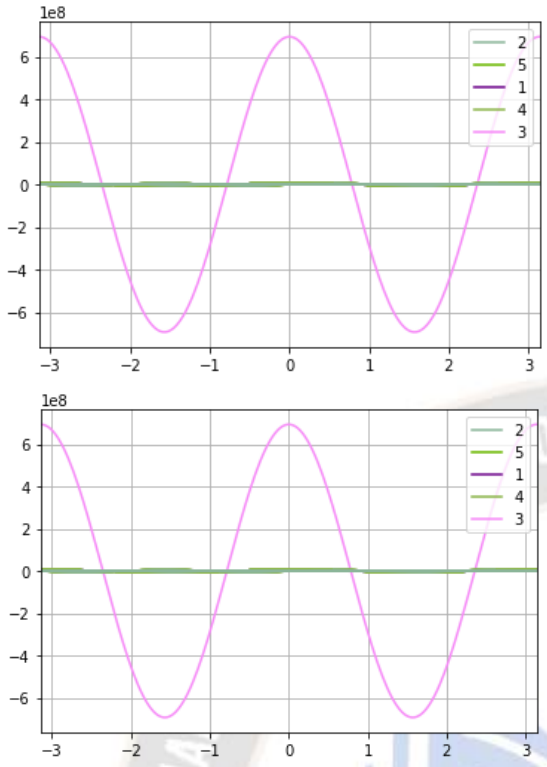


Fig. 8. Andrew curve for KDD-41and Normalized KDD-41

Andrew Curves For CICIDS-2017 Dataset

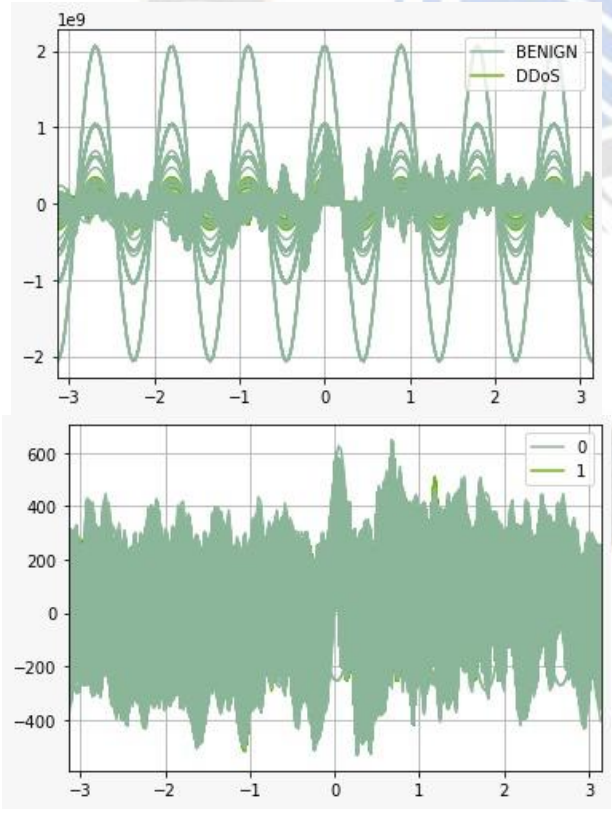


Fig. 9. Andrew curve for CICIDS-2017 and Normalized CICIDS-2017

From Fig 4 to Fig 9, Andrew Curves applied on instances using the datasets. Finally, after seeing the Andrew curves of all the datasets we can conclude that the nature of the datasets is highly non-linear.

Csv Files to Gray Scale Images

By using the Intrusion Detection Evaluation Dataset (CIC-IDS2019) we have converted the csv files to gray scale image dataset. It will be useful if the model takes images as input to detect the attack. There are mainly consists of 14 classes. The below images are one of the different class datasets. The images are generated by dropping null, outliers and infinity rows, normalizing the dataset, converting the dataset into integer type and multiplying it with 255. These dataset images are given as input to the proposed model. We have also uploaded this image dataset in IEEE data port, the link given as Link: https://ieee-dataport.org/documents/cloud-attack-dataset.

V. RESULTS AND DISCUSSION

In this paper, we have contributed to machine learning, and several techniques have been implemented. There are six algorithms used in this work, and they are Navie Bayes, KNN and SVM with FFSC and without FFSC.

Result with Feature Feature Score (FFSC):

Detailed description about the Train data set

- Number of traffic instances: 27686
- Number of features: 63
- Class labels with with number of traffic instances in each:
  Attack -> 25186
  Benign -> 2500

Detailed description about the Test data set

- Number of traffic instances: 12622
- Number of features: 63
- Class labels with with number of traffic instances in each:
  Attack -> 11606
  Benign -> 1016

Generated Andrew curves:

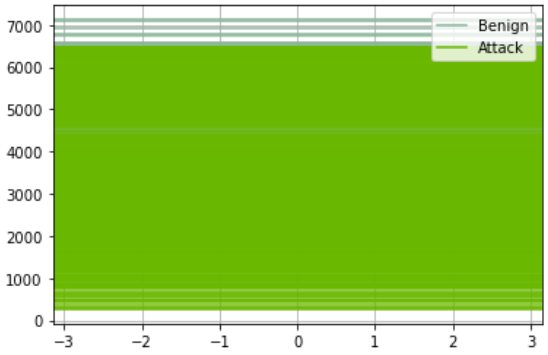


Fig. 10. Andrew curves generated with FFSC dataset

Naive Bayes Results: Train dataset for: confusion matrix:

[[25186 0]
[2500 0]]

Overall result:

- Accuracy -> 90.97016542656938

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
Article Received: 09 April 2023 Revised: 30 May 2023 Accepted: 12 June 2023

- Precision -> 90.97016542656938
- Recall -> 90.97016542656938
- F-score -> 90.97016542656938

Class	Sensitivity	Specificity	Balanced ac curacy
Attack	0.0	1.0	0.5
Benign	1.0	0.0	0.5

Classification Report

Class	Precision	Recall	F-1 Score	Support
Attack	0.9	1.00	0.95	25186
Benign	0.00	0.00	0.00	2500

ROC-AUC Curve:

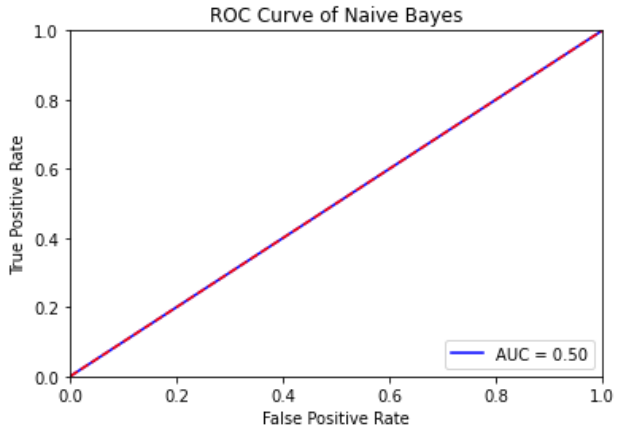


Fig. 11. ROC curve for Naïve Bayes for Train data set

Test dataset: confusion matrix:

[[11606 0]
[1016 0]]

Overall result:

- Accuracy -> 91.95056250990335
- Precision -> 91.95056250990335
- Recall -> 91.95056250990335
- F-score -> 91.95056250990335

Class	Sensitivity	Specificity	Balanced acc uracy
Attack	0.0	1.0	0.5
Benign	1.0	0.0	0.5

Classification Report

Class	Precision	Recall	F-1 Score	Support
Attack	0.91	1.00	0.95	25186
Benign	0.00	0.00	0.00	2500

ROC-AUC Curve:

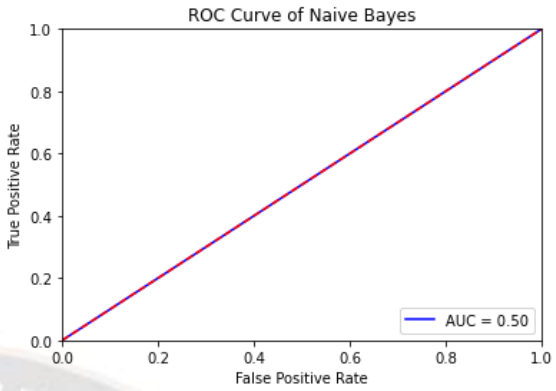


Fig. 12. ROC curve for Naïve Bayes for Test data set

K Neighbors Classifier Results: n value -> 1

Train dataset for: confusion matrix:

[[25186 0]
[0 2500]]

Overall result:

- Accuracy -> 100.0
- Precision -> 100.0
- Recall -> 100.0
- F-score -> 100.0

Classification Report:

Class	Sensitivity	Specificity	Balanced acc uracy
Attack	0.0	1.0	1.0
Benign	1.0	1.0	1.0

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	1.00	1.00	25186
Benign	1.00	1.00	1.00	2500

ROC-AUC Curve:

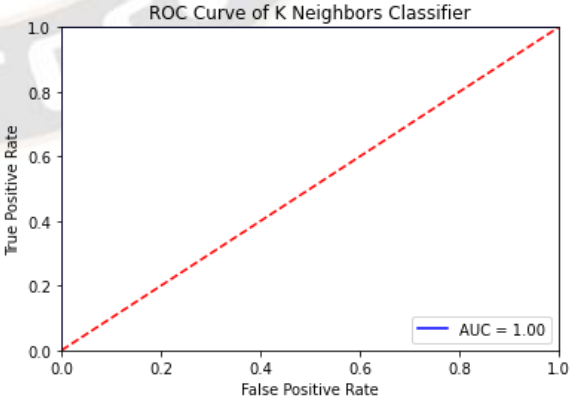


Fig. 13. ROC curve for K Neighbors for Train data set

Test dataset: confusion matrix:

[[10920 686]
[914 102]]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

Overall result:

- Accuracy -> 87.32372048803676
- Precision -> 87.32372048803676
- Recall -> 87.32372048803676
- F-score -> 87.32372048803676

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.100394	0.940893	0.520643
Benign	0.940893	0.100394	0.520643

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	1.00	1.00	25186
Benign	1.00	1.00	1.00	2500

ROC-AUC Curve:

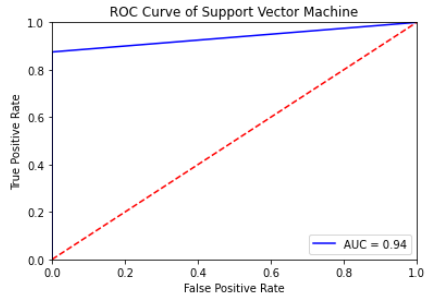


Fig. 14. ROC curve for Support Vector Machine for Train data set

Test dataset: confusion matrix:

[[11111 495]
[941 75]]

Overall result:

- Accuracy -> 88.623039138013
- Precision -> 88.623039138013
- Recall -> 88.623039138013
- F-score -> 88.623039138013

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.07819	0.957350	0.515584
Benign	0.957350	0.073819	0.515584

Class	Precision	Recall	F-1 Score	Support
Attack	0.98	1.00	0.99	25186
Benign	0.95	0.83	0.89	2500

ROC-AUC Curve:

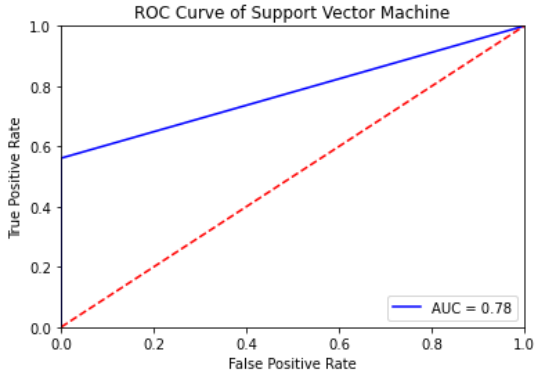


Fig. 15. ROC curve for Support Vector Machine for Test data set

Decision Tree Entropy Results: Train dataset for: confusion matrix:

[[24976 210]
[820 1680]]

Overall result:

- Accuracy -> 96.27970815574659
- Precision -> 96.27970815574659
- Recall -> 96.27970815574659
- F-score -> 96.27970815574659

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.672000	0.991662	0.831831
Benign	0.991662	0.672000	0.831831

Class	Precision	Recall	F-1 Score	Support
Attack	0.97	0.99	0.98	25186
Benign	0.89	0.67	0.77	2500

ROC-AUC Curve:

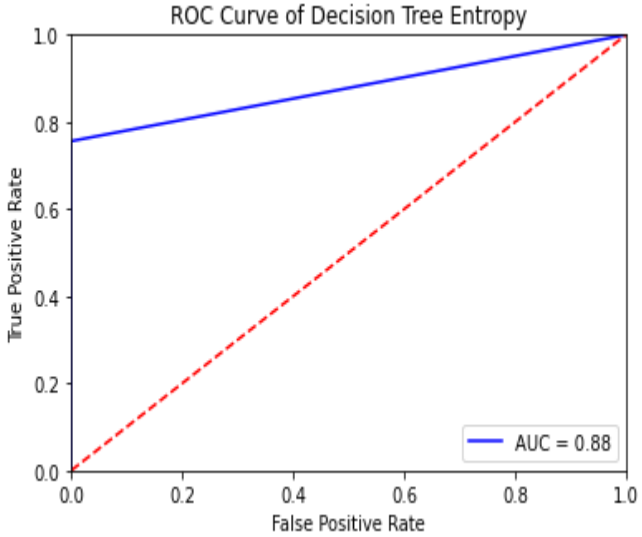


Fig. 16. ROC curve for Decision Tree for Train data set

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
Article Received: 09 April 2023 Revised: 30 May 2023 Accepted: 12 June 2023

Test dataset: confusion matrix:

[[11302 304]
[1008 8]]

Overall result:

- Accuracy -> 89.60545080019014
- Precision -> 89.60545080019014
- Recall -> 89.60545080019014
- F-score -> 89.60545080019014

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.007874	0.973807	0.49084
Benign	0.973807	0.007874	0.49084

Class	Precision	Recall	F-1 Score	Support
Attack	0.97	0.99	0.98	25186
Benign	0.89	0.67	0.77	2500

ROC-AUC Curve:

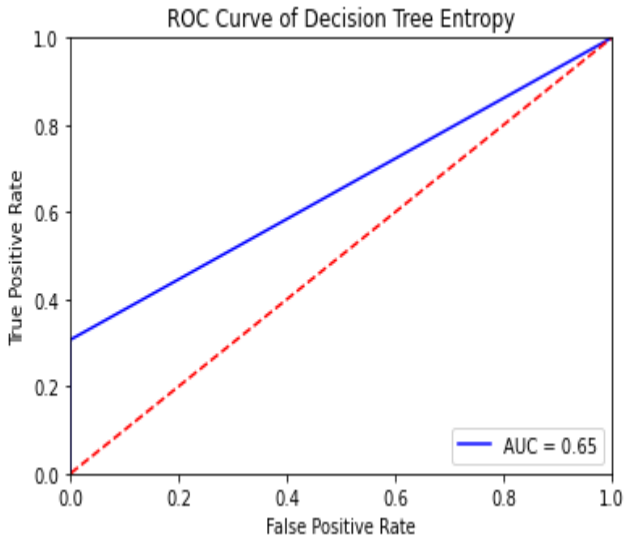


Fig. 17. ROC curve for Decision Tree entropy for Test data set

Result with Out Feature Feature Score (FFSC):

Detailed description about the Train data set:

- Number of traffic instances: 27686
- Number of features: 63
- Class labels with with number of traffic instances in each:
  Attack -> 25186
  Benign -> 2500

Detailed description about Test Dataset:

- Number of traffic instances: 12622
- Number of features: 63
- Class labels with with number of traffic instances in each:
  Attack -> 11606
  Benign -> 1016

Generated Andrew curves without FFSC:

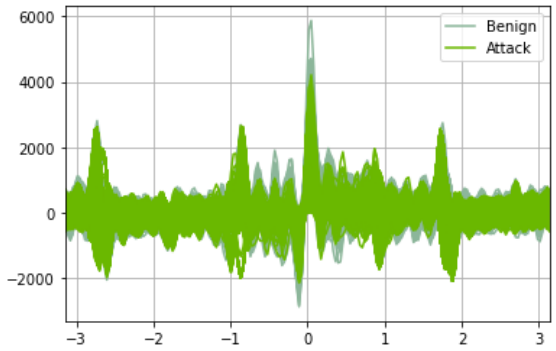


Fig. 18. Generated Andrew curves without FFSC

Naive Bayes Results:

Train dataset for:

confusion matrix:

[[22420 2766]
[23 2477]]

Overall result:

- Accuracy -> 89.92631654988081
- Precision -> 89.92631654988081
- Recall -> 89.92631654988081
- F-score -> 89.92631654988081

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.990800	0.890177	0.940489
Benign	0.890177	0.990800	0.040489

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	0.89	0.94	25186
Benign	0.47	0.99	0.64	2500

ROC-AUC Curve:

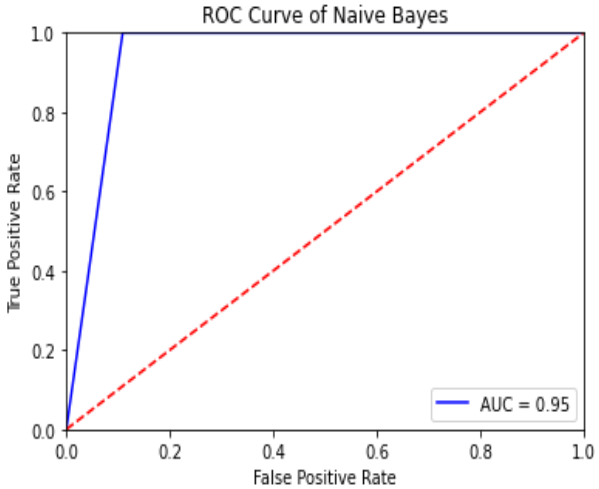


Fig. 19. ROC curve of Navie Bayes for train data set with out FFSC

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
Article Received: 09 April 2023 Revised: 30 May 2023 Accepted: 12 June 2023

Test dataset:

confusion matrix:

[[11572 34]
[1016 0]]

Overall result:

- Accuracy -> 91.68119157027412
- Precision -> 91.68119157027412
- Recall -> 91.68119157027412
- F-score -> 91.68119157027411

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.00000	0.99707	0.498535
Benign	0.99707	0.00000	0.498535

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	0.89	0.94	25186
Benign	0.47	0.99	0.64	2500

ROC-AUC Curve:

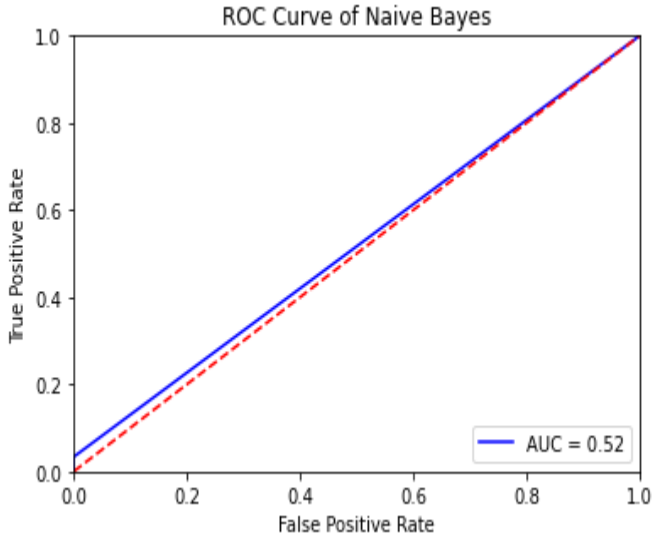


Fig. 20. ROC curve of Navie Bayes for test data set with out FFSC

K Neighbors Classifier Results: n value -> 1

Train dataset for: confusion matrix:

[[25186 0]
[0 2500]]

Overall result:

- Accuracy -> 100.0
- Precision -> 100.0
- Recall -> 100.0
- F-score -> 100.0

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	1.0	1.0	1.0
Benign	1.0	1.0	1.0

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	1.00	1.00	25186
Benign	1.00	1.00	1.00	2500

ROC-AUC Curve:

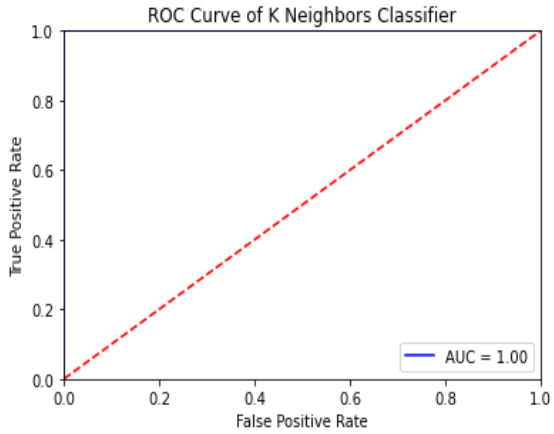


Fig. 21. ROC curve of K Neighbors Classifier for train data set with out FFSC

Test dataset: confusion matrix:

[[11606 0]
[1016 0]]

Overall result:

- Accuracy -> 91.95056250990335
- Precision -> 91.95056250990335
- Recall -> 91.95056250990335
- F-score -> 91.95056250990335

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.0	1.0	0.5
Benign	1.0	1.0	0.5

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	1.00	1.00	25186
Benign	1.00	1.00	1.00	2500

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
Article Received: 09 April 2023 Revised: 30 May 2023 Accepted: 12 June 2023

ROC-AUC Curve:

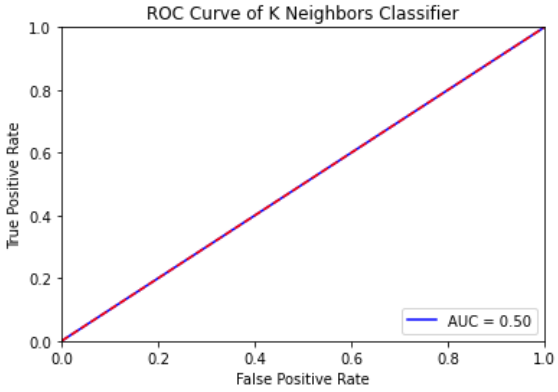


Fig. 22. ROC curve of K Neighbors Classifier for test data set with out FFSC

Support Vector Machine Results: Kernel selected -> rbf
Train dataset for: confusion matrix:

[[25186 0]
[0 2500]]

Overall result:

- Accuracy -> 100.0
- Precision -> 100.0
- Recall -> 100.0
- F-score -> 100.0

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.0	1.0	1.0
Benign	1.0	1.0	1.0

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	1.00	1.00	25186
Benign	1.00	1.00	1.00	2500

ROC-AUC Curve:

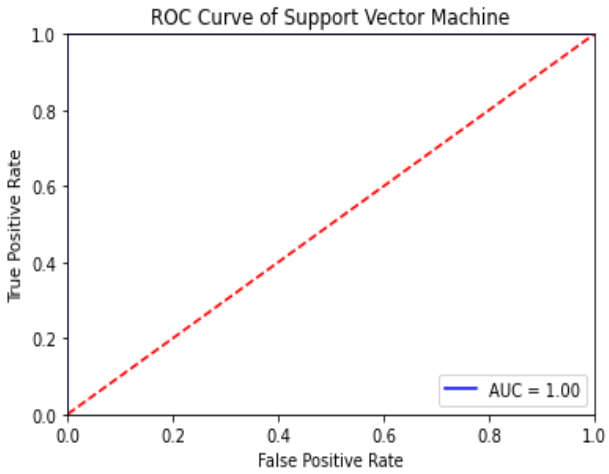


Fig. 23. ROC curve of Support Vector Machine for train data set with out FFSC

Test dataset: confusion matrix:

[[11606 0]
[1016 0]]

Overall result:

- Accuracy -> 91.95056250990335
- Precision -> 91.95056250990335
- Recall -> 91.95056250990335
- F-score -> 91.95056250990335

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.0	1.0	0.5
Benign	1.0	1.0	0.5

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	1.00	1.00	25186
Benign	1.00	1.00	1.00	2500

ROC-AUC Curve:

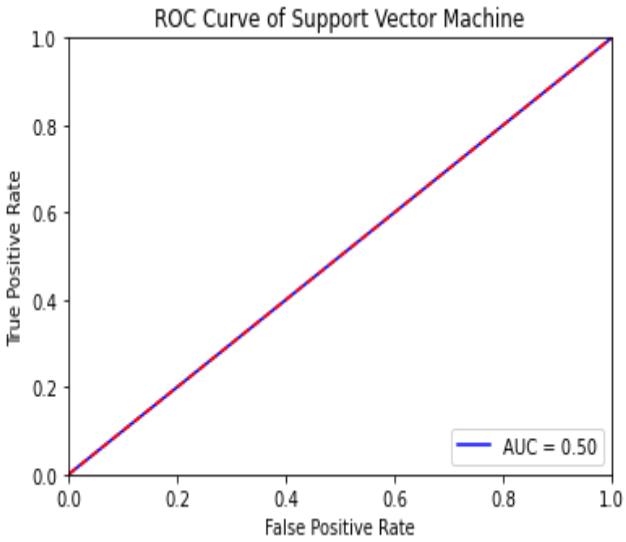


Fig. 24. ROC curve of Support Vector Machine for test data set with out FFSC

Decision Tree Entropy Results:
Train dataset for: confusion matrix:

[[25186 0]
[0 2500]]

Overall result:

- Accuracy -> 100.0
- Precision -> 100.0
- Recall -> 100.0
- F-score -> 100.0

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.7048
***Article Received**: 09 April 2023 **Revised:** 30 May 2023 **Accepted:** 12 June 2023*

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.0	1.0	1.0
Benign	1.0	1.0	1.0

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	1.00	1.00	25186
Benign	1.00	1.00	1.00	2500

ROC-AUC Curve:

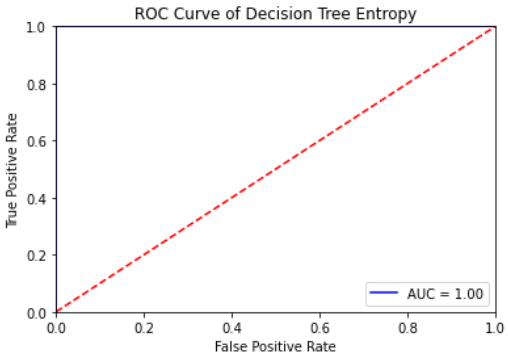


Fig. 25. ROC curve of Decision Tree for train data set with out FFSC

Test dataset: confusion matrix:

[[10499 1107]
[1016 0]]

Overall result:

- Accuracy -> 83.18016162256377
- Precision -> 83.18016162256377
- Recall -> 83.18016162256377
- F-score -> 83.18016162256379

Classification Report:

Class	Sensitivity	Specificity	Balanced accuracy
Attack	0.0000000	0.904618	0.452309
Benign	0.904618	0.000000	0.452309

Class	Precision	Recall	F-1 Score	Support
Attack	1.00	1.00	1.00	25186
Benign	1.00	1.00	1.00	2500

ROC-AUC Curve:

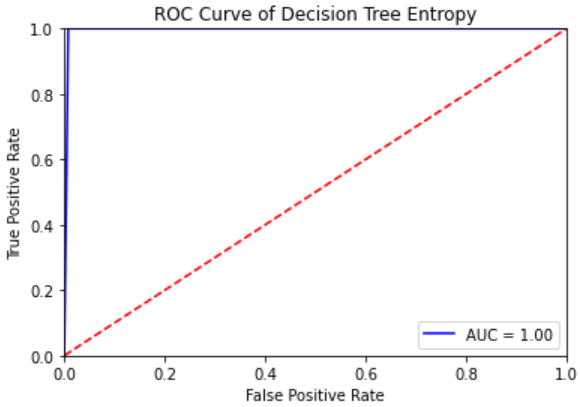


Fig. 26. ROC curve of Decision Tree for train data set with out FFSC

VI. CONCLUSION

The main thought of the proposed strategy is to distinguish low rate DDoS assault successfully by utilizing a few stages like pre-handling, include choice and order.There are a variety of subclassified attacks that must also be correctly identified.The proposed method ought to provide the highest levels of precision, recall, and F-measure accuracy.For better comprehension of the system's operation, we would like to implement a web or mobile application to demonstrate the learning model's implemented functionality.utilizing a variety of tools and technologies like Flask, Nodejs, HTML, CSS, Javascript, Angular, Bootstrap, Database.

REFERENCES

[1] Frederico A. F. Silveira,Agostinho De Medeiros Brito Junior,Genoveva Vargas-Solar, And Luiz F. Silveira, “Smart Detection: An Online Approach For Dos/Ddos Attack Detection Using Machine Learning”,Volume 2019 |Article Id 1574749

[2] Ivandro Ortet Lopes, Deqing Zou,Francis A Ruambo,Saeed Akbar, And Bin Yuan, “Towards Effective Detection Of Recent Ddos Attacks: A Deep Learning Approach”, Volume 2021 |Article Id 5710028

[3] P. Renuka ,Dr. B. Booba, Professor ,”Analysis On Detecting Ddos Attack In Iot Environment “,2018, Issn : 0731-6755

[4] Yuanyuan Wei; Julian Jang-Jaccard; Fariza Sabrina; Amardeep Singh; Wen Xu; Seyit Camtepe,”Ae-Mlp: A Hybrid Deep Learning Approach For Ddos Detection And Classification”,Ieee Access (Volume: 9)

[5] Zecheng He; Tianwei Zhang; Ruby B. Lee,”Machine Learning Based Ddos Attack Detection From Source Side In Cloud”,2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)

[6] B. Narsimha, Ch V Raghavendran, Pannangi Rajyalakshmi, G Kasi Reddy, M. Bhargavi and P. Naresh (2022), Cyber Defense in the Age of Artificial Intelligence and Machine Learning for Financial Fraud Detection Application. IJEER 10(2), 87-92. DOI: 10.37391/IJEER.100206.

[7] Naresh, P., & Suguna, R. (2021). IPOC: An efficient approach for dynamic association rule generation using incremental data with updating supports. Indonesian Journal of Electrical Engineering and Computer Science, 24(2), 1084. https://doi.org/10.11591/ijeecs.v24.i2.pp1084-1090.

[8] Mark White, Thomas Wood, Carlos Rodríguez, Pekka Koskinen, Jónsson Ólafur. Exploring Natural Language Processing in Educational Applications. Kuwait Journal of Machine Learning, 2(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/168

[9] Kolawole Abubakar Sadiq, Aderonke Thompson, "Mitigating DDoS Attacks in Cloud Network using Fog and SDN: A Conceptual Security Framework", DOI:10.5120/ijais2020451877, August 2020

[10] Swathi Sambangi,Lakshmeeswari Gondi, "A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression", 25 December 2020.

[11] Er. Sakshi kakkar , Er. Dinesh kumar, "A Survey on Distributed Denial of Services (DDOS) ",Vol.5(3), 2014,3863-3866.

[12] Nazrul Hoque; Dhruba K Bhattacharyya; Jugal K Kalita,"A novel measure for low-rate and high-rate DDoS attack detection using multivariate data analysis", 2016 8th International Conference on Communication Systems and Networks.

[13] Ankit Agarwal, Manju Khari & Rajiv Singh, "Detection of DDOS Attack using Deep Learning Model in Cloud Storage Application", Wireless Pers Commun (2021).

[14] T. Aruna, P. Naresh, A. Rajeshwari, M. I. T. Hussan and K. G. Guptha, "Visualization and Prediction of Rainfall Using Deep Learning and Machine Learning Techniques," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 910-914, doi: 10.1109/ICTACS56270.2022.9988553.

[15] V. Krishna, Y. D. Solomon Raju, C. V. Raghavendran, P. Naresh and A. Rajesh, "Identification of Nutritional Deficiencies in Crops Using Machine Learning and Image Processing Techniques," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 925-929, doi: 10.1109/ICIEM54221.2022.9853072.

[16] Naresh, P., & Suguna, R. (2019). Association Rule Mining Algorithms on Large and Small Datasets: A Comparative Study. 2019 International Conference on Intelligent Computing and Control Systems (ICCS). DOI:10.1109/iccs45141.2019.9065836.

[17] Naresh, K. Pavan kumar, and D. K. Shareef, 'Implementation of Secure Ranked Keyword Search by Using RSSE,' International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 2 Issue 3, March – 2013.

[18] S, D. A. (2021). CCT Analysis and Effectiveness in e-Business Environment. International Journal of New Practices in Management and Engineering, 10(01), 16–18. https://doi.org/10.17762/ijnpme.v10i01.97

[19] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh (2022), Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. IJEER 10(2), 80-86. DOI: 10.37391/IJEER.100205.. https://doi.org/10.18280/ria.360107.

[20] S. Khaleelullah, P. Marry, P. Naresh, P. Srilatha, G. Sirisha and C. Nagesh, "A Framework for Design and Development of Message sharing using Open-Source Software," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 639-646, doi: 10.1109/ICSCDS56580.2023.10104679.

[21] Dasari, K.B., Devarakonda, N. (2021). Detection of different DDoS attacks using machine learning classification algorithms. Ingénierie des Systèmes d'Information, Vol. 26, No. 5, pp. 461-468. https://https://doi.org/10.18280/isi.260505.