A Classification and Prediction Method for DDoS Attacks Based on Machine Learning

B.Raghupathi[1]|Dr.P.Satish Reddy[2]|Asif Ahmed Algur[3]|KARNE ANUSHKA[4]

1, 2 & 3 Associate Professor, CSE department, Kasireddy Narayanreddy College of Engineering And Research, Hyderabad, TS.

4 UG SCHOLAR, CSE department, Kasireddy Narayanreddy College of Engineering And Research, Hyderabad, TS.

ABSTRACT: Distributed Denial of Service (DDoS) assaults are the most common term used to describe distributed network attacks. These assaults exploit certain restrictions that are applicable to each arrangement asset, including the structure of the website of the authorized organization. The author worked on an outdated KDD dataset for the current study. To determine the current state of DDoS attacks, the most recent dataset must be used. This study classified and predicted the types of DDoS attacks using a machine learning technique. XGBoost and Random Forest classification algorithms were employed for this. A comprehensive methodology for predicting DDoS assaults was proposed in order to access the research. Python was utilized as a simulator for the proposed work, and the UNWS-np-15 dataset was taken from the GitHub source. We created a confusion matrix to identify the model performance after applying the machine learning models. The Random Forest algorithm's precision (PR) and recall (RE) in the initial classification were both 89%, according to the results. Our suggested model's average Accuracy (AC) is 89%, which is excellent and sufficient. According to the results, the XG Boost algorithm's Precision (PR) and Recall (RE) are both around 90% in the second classification. Our proposed model's average Accuracy (AC) is 90%. The accuracy of the fault identification was much increased by comparing our work to previous research studies, which are roughly 85% and 79%, respectively.

KEYWORDS: AC, DDOS, XGBOOST, GitHUB, UNWS.

I.INTRODUCTION: Distributed network attacks are referred to, usually, as Distributed Denial of Service (DDOS) attacks. These attacks take advantage of specific limitations that apply to any arrangement asset, such as the framework of the authorized organization's website. A DDOS attack sends different requests (with IP spoo_ng) to the target web assets to exceed the site's ability to handle various requests, at a given time, and make the site unable to operate effectively and efficiently _ even for the legitimate users of the network. Typically, the target of various DDOS attacks are web applications and business websites; and the attacker may have different goals [1], [2]. Some common types of the DDOS attacks are shown in Figure 1. We give brief description of each attack in Section I-A. The Internet of Things (IOT) implies the arrangement of interconnected, web-related objects that can collect and interchange information through remote organizations without manual intervention [3]. The ``Things" can simply be related clinical tools, bio-chip transponders, solar panels, and related vehicles with sensors that can warn the driver of numerous potential problems [4], or any article with sensors that can collect and move information in the organization. Artificial intelligence (AI) is a small tool that transforms information into data. In the past 50 years (approximately), information has had an impact on users privacy and security. Except for the possibility of researching it and finding the examples hidden in it, the amount of information is negligible. Artificial intelligence technology is usually used to find important secret examples in complex information, and this work will try to find them in some way. Mysterious examples and data about a problem can be used to predict future events and play a wide range of complex dynamics. There were different approaches proposed for DDOS attack classification and prevention.

In [4] deep learning models are proposed for intrusion detection. The dataset was UNSW-nb15 and the models were Convention neural network (CNN), BAT-MC, BAT, and Recurrent neural network. The overall model's performance was very good. They found CNN best for the proposal. The average accuracy was 79%. In paper [5] authors proposed a hybrid model deep learning model for intrusion detection. They combined two deep learning for the classification of CNN and LSTM from the RNN model. The dataset was used in this work is KDD. They found an 85.14% average accuracy for the proposed. However, up to our knowledge different deep learning models are used for DDOS attacks. Similarly, they used the same KDD dataset from the UCI repository in research. In finally all authors found the same results 85%.

II.EXISTING SYSTEM: We studied the latest research papers of the past two years for this research work and also Gozde

Karatas et al. [2] proposed a machine learning approach for attacks classification. They used different machine learning algorithms and found that the KNN model is best for classification as compared to other research work. Nuno Martins et al. [1] proposed intrusion detection using machine learning approaches. They used the KDD dataset which is available on the UCI repository. They performed different supervised models to balance un classification algorithm for better performance. In this work, a comparative study was proposed by the use of different classification algorithms and found good results in their work.

III.PROPOSED SYSTEM: In this research, we design a framework for the DDoS attack classification and prediction based on the existing dataset that used machine learning methods. This framework involves the following main steps.

1) The first step involves the selection of dataset for utilization.

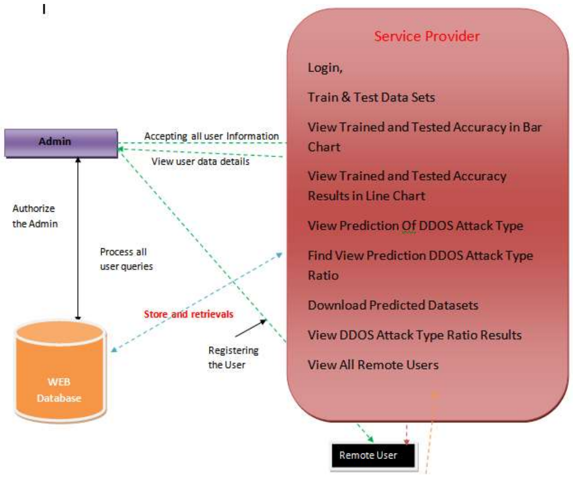
2) The second step involves the selection of tools and language.

3) The third step involves data pre-processing techniques to handle irrelevant data from the dataset. In the fourth step feature extraction and label.

4) Encoding is performed to convert symbolical data into numerical data.

5) In the fifth step, the data splitting is performed into a train and test set for the model.

In this step, we build and train our proposed model. However, model optimization is also performed on the trained model in terms of kernel scaling and kernel hyper-parameter tuning to improve model efficiency. When the model optimizes then we will generate output results from the model.


Service Provider
Login,
Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results in Line Chart
View Prediction Of DDOS Attack Type
Find View Prediction DDOS Attack Type Ratio
Download Predicted Datasets
View DDOS Attack Type Ratio Results
View All Remote Users
Admin
Accepting all user Information
View user data details
Authorize the Admin
Process all user queries
Store and retrievals
Registering the User
WEB Database
Remote User

3.1 Architecture Diagram

The main contribution is to generate the best model for data utilization, as well as, model optimization; and which performs best for model learning. After getting the results, we performed performance measures in terms of precision, recall, and f1 score. In this research work, we used two well known supervised learning models which are: (i) Random Forest Classifier; and (ii) XGBoost

Classifier. The architecture and data flow diagram of the proposed method is shown in Figure 3.2.

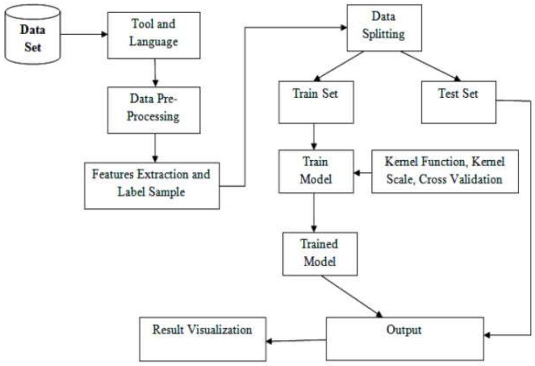
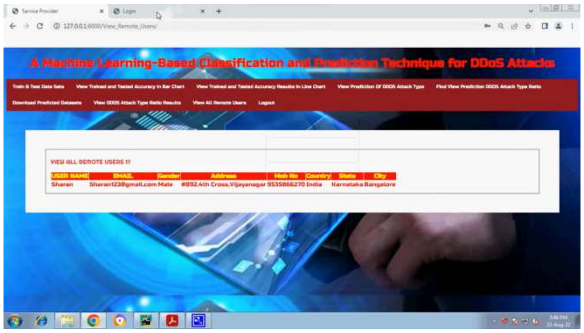


FIGURE 3.2. Data flow chart for the proposed machine learning based DDoD attack prediction technique.

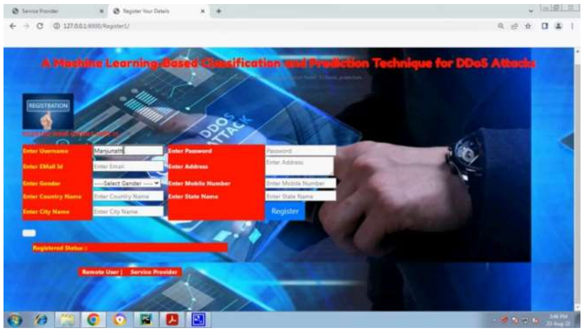
IV.RESULTS: We selected the UNSW-nb15 dataset from GitHub1 that contains features' data about the DDoS attacks. This dataset is provided by the Australian Centre for Cyber Security (ACCS) [25]. Table 1 shows the total numbers of rows and columns in the dataset. The dataset consists of different features about the DDoS attacks including an ID number, Proto which presents medium of the network, label of the attacks, and attacks' cat which presents the severity of the DDoS attacks.


Service Provider
Login
127.0.0.1:8000/View_Remote_Users/
Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results in Line Chart
View Prediction Of DDOS Attack Type
Find View Prediction DDOS Attack Type Ratio
Download Predicted Datasets
View DDOS Attack Type Ratio Results
View All Remote Users
Logout

A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks

VIEW ALL REMOTE USERS !!!

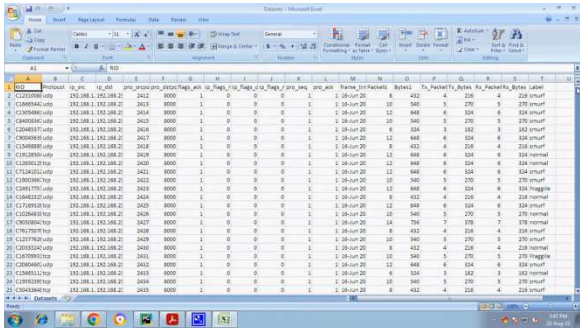
USER NAME	EMAIL	Gender	Address	Mob No	Country	State	City
Sharan	Sharan123@gmail.com	Male	#892,4th Cross,Vijayanagar	9535866270	India	Karnataka	Bangalore



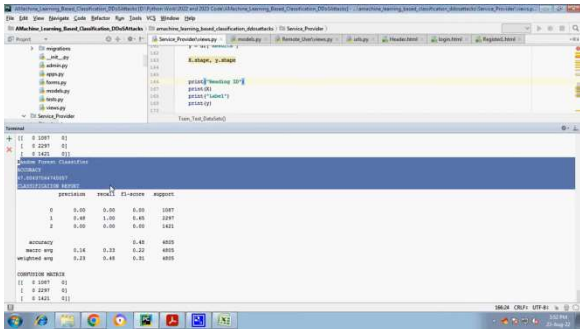
A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks

Enter Username: Manjunath
Enter EMail Id: Enter Email
Enter Gender: ----Select Gender ----
Enter Country Name: Enter Country Name
Enter City Name: Enter City Name
Enter Password: Password
Enter Address: Enter Address
Enter Mobile Number: Enter Mobile Number
Enter State Name: Enter State Name

Registered Status :

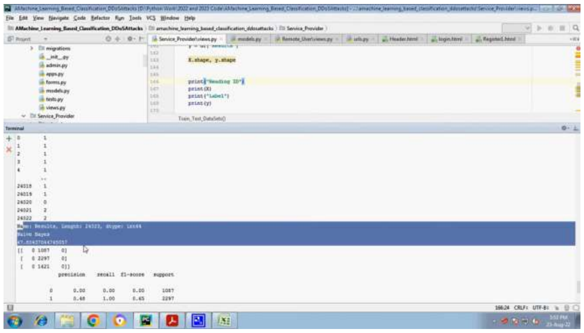


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	RID	Protocol	ip_src	ip_dst	pro_srcpo	pro_dstpc	flags_ack	ip_flags_r	ip_flags_c	ip_flags_r	pro_seq	pro_ack	frame_tin	Packets	Bytes1	Tx_Packet	Tx_Bytes	Rx_Packet	Rx_Bytes	Label
2	C1231006	udp	192.168.1.	192.168.2	2412	8000	1	0	0	0	1	1	16-Jun-20	8	432	4	216	4	216	smurf
3	C1666544	udp	192.168.1.	192.168.2	2413	8000	1	0	0	0	1	1	16-Jun-20	10	540	5	270	5	270	smurf
4	C1305486	udp	192.168.1.	192.168.2	2414	8000	1	0	0	0	1	1	16-Jun-20	12	648	6	324	6	324	smurf
5	C8400836	udp	192.168.1.	192.168.2	2415	8000	1	0	0	0	1	1	16-Jun-20	10	540	5	270	5	270	smurf
6	C2048537	udp	192.168.1.	192.168.2	2416	8000	1	0	0	0	1	1	16-Jun-20	6	324	3	162	3	162	smurf
7	C9004563	udp	192.168.1.	192.168.2	2417	8000	1	0	0	0	1	1	16-Jun-20	12	648	6	324	6	324	smurf
8	C1549888	udp	192.168.1.	192.168.2	2418	8000	1	0	0	0	1	1	16-Jun-20	8	432	4	216	4	216	smurf
9	C1912850	udp	192.168.1.	192.168.2	2419	8000	1	0	0	0	1	1	16-Jun-20	12	648	6	324	6	324	normal
10	C1265012	tcp	192.168.1.	192.168.2	2420	8000	1	0	0	0	1	1	16-Jun-20	12	648	6	324	6	324	normal
11	C7124101	udp	192.168.1.	192.168.2	2421	8000	1	0	0	0	1	1	16-Jun-20	12	648	6	324	6	324	smurf
12	C1900366	tcp	192.168.1.	192.168.2	2422	8000	1	0	0	0	1	1	16-Jun-20	10	540	5	270	5	270	smurf
13	C2491775	udp	192.168.1.	192.168.2	2423	8000	1	0	0	0	1	1	16-Jun-20	12	648	6	324	6	324	Fraggile
14	C1648232	udp	192.168.1.	192.168.2	2424	8000	1	0	0	0	1	1	16-Jun-20	8	432	4	216	4	216	normal
15	C1716932	tcp	192.168.1.	192.168.2	2425	8000	1	0	0	0	1	1	16-Jun-20	12	648	6	324	6	324	smurf
16	C1026483	tcp	192.168.1.	192.168.2	2426	8000	1	0	0	0	1	1	16-Jun-20	10	540	5	270	5	270	normal
17	C9050804	tcp	192.168.1.	192.168.2	2427	8000	1	0	0	0	1	1	16-Jun-20	14	756	7	378	7	378	normal
18	C7617507	tcp	192.168.1.	192.168.2	2428	8000	1	0	0	0	1	1	16-Jun-20	8	432	4	216	4	216	smurf
19	C1237762	udp	192.168.1.	192.168.2	2429	8000	1	0	0	0	1	1	16-Jun-20	10	540	5	270	5	270	smurf
20	C2033524	udp	192.168.1.	192.168.2	2430	8000	1	0	0	0	1	1	16-Jun-20	8	432	4	216	4	216	normal
21	C1670993	tcp	192.168.1.	192.168.2	2431	8000	1	0	0	0	1	1	16-Jun-20	10	540	5	270	5	270	Fraggile
22	C2080460	udp	192.168.1.	192.168.2	2432	8000	1	0	0	0	1	1	16-Jun-20	12	648	6	324	6	324	smurf
23	C1566511	tcp	192.168.1.	192.168.2	2433	8000	1	0	0	0	1	1	16-Jun-20	6	324	3	162	3	162	normal
24	C1959239	tcp	192.168.1.	192.168.2	2434	8000	1	0	0	0	1	1	16-Jun-20	10	540	5	270	5	270	smurf
25	C5043384	tcp	192.168.1.	192.168.2	2435	8000	1	0	0	0	1	1	16-Jun-20	8	432	4	216	4	216	smurf



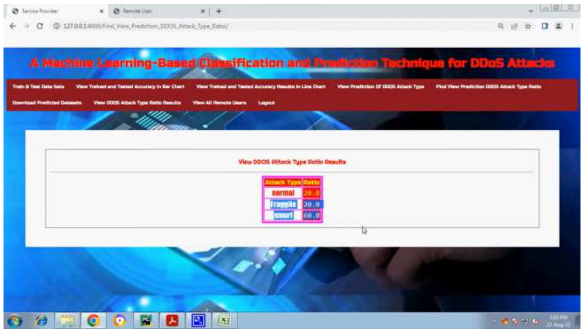
```
X.shape, y.shape


print("Reading ID")
print(X)
print("Label")
print(y)
```

```
[[   0 1087    0]
 [   0 2297    0]
 [   0 1421    0]]
Random Forest Classifier
ACCURACY
47.80437044745057
CLASSIFICATION REPORT
              precision    recall  f1-score   support

           0       0.00      0.00      0.00      1087
           1       0.48      1.00      0.65      2297
           2       0.00      0.00      0.00      1421

    accuracy                           0.48      4805
   macro avg       0.16      0.33      0.22      4805
weighted avg       0.23      0.48      0.31      4805

CONFUSION MATRIX
[[   0 1087    0]
 [   0 2297    0]
 [   0 1421    0]]
```



```
X.shape, y.shape


print("Reading ID")
print(X)
print("Label")
print(y)
```

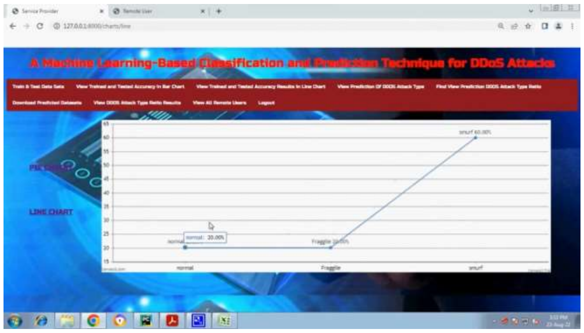
```
0        1
1        1
2        1
3        1
4        1
        ..
24018    1
24019    1
24020    0
24021    2
24022    2
Name: Results, Length: 24023, dtype: int64
Naive Bayes
47.81437044745057
[[   0 1087    0]
 [   0 2297    0]
 [   0 1421    0]]
              precision    recall  f1-score   support

           0       0.00      0.00      0.00      1087
           1       0.48      1.00      0.65      2297
```

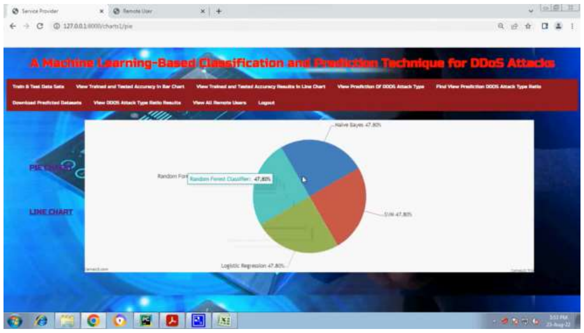

Service Provider
Remote User
127.0.0.1:8000/Find_View_Prediction_DDOS_Attack_Type_Ratio/
A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks
Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results in Line Chart
View Prediction Of DDOS Attack Type
Find View Prediction DDOS Attack Type Ratio
Download Predicted Datasets
View DDOS Attack Type Ratio Results
View All Remote Users
Logout

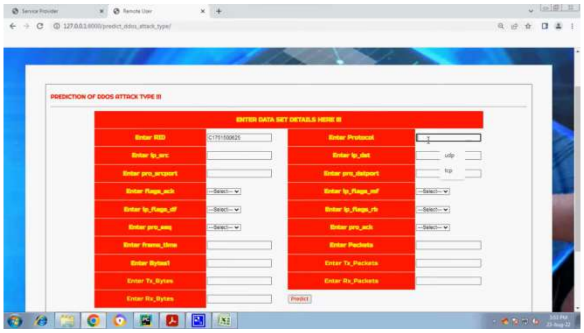
View DDOS Attack Type Ratio Results

Attack Type	Ratio
normal	20.0
Fraggile	20.0
smurf	60.0

Service Provider
Remote User
127.0.0.1:8000/likeschart/bar
A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks
Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results in Line Chart
View Prediction Of DDOS Attack Type
Find View Prediction DDOS Attack Type Ratio
Download Predicted Datasets
View DDOS Attack Type Ratio Results
View All Remote Users
Logout
DDOS ATTACK


Service Provider
Remote User
127.0.0.1:8000/charts/line
A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks
Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results in Line Chart
View Prediction Of DDOS Attack Type
Find View Prediction DDOS Attack Type Ratio
Download Predicted Datasets
View DDOS Attack Type Ratio Results
View All Remote Users
Logout
PIE CHART
LINE CHART
65
60
55
50
45
40
35
30
25
20
15
normal: 20.00%
Fraggile 20.00%
smurf 60.00%
normal
Fraggile
smurf

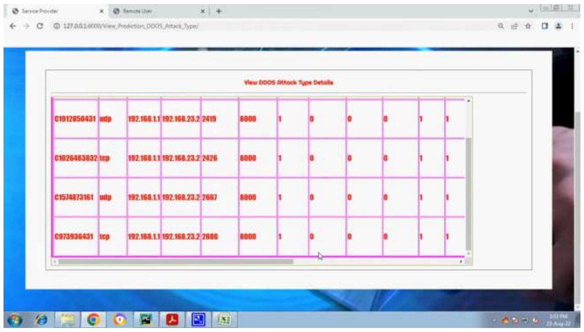

Service Provider
Remote User
127.0.0.1:8000/charts1/pie
A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks
Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results in Line Chart
View Prediction Of DDOS Attack Type
Find View Prediction DDOS Attack Type Ratio
Download Predicted Datasets
View DDOS Attack Type Ratio Results
View All Remote Users
Logout
PIE CHART
LINE CHART
Naive Bayes 47.80%
Random Forest Classifier: 47.80%
SVM 47.80%
Logistic Regression 47.80%



PREDICTION OF DDOS ATTACK TYPE !!!

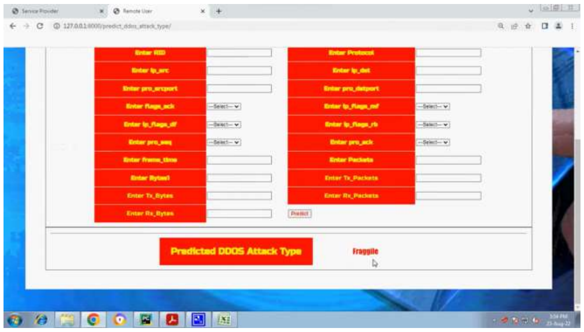
ENTER DATA SET DETAILS HERE !!!

Enter RID	C1751500625	Enter Protocol	
Enter ip_src		Enter ip_dst	
Enter pro_srcport		Enter pro_dstport	
Enter flags_ack	--Select--	Enter ip_flags_mf	--Select--
Enter ip_flags_df	--Select--	Enter ip_flags_rb	--Select--
Enter pro_seq	--Select--	Enter pro_ack	--Select--
Enter frame_time		Enter Packets	
Enter Bytes1		Enter Tx_Packets	
Enter Tx_Bytes		Enter Rx_Packets	
Enter Rx_Bytes			



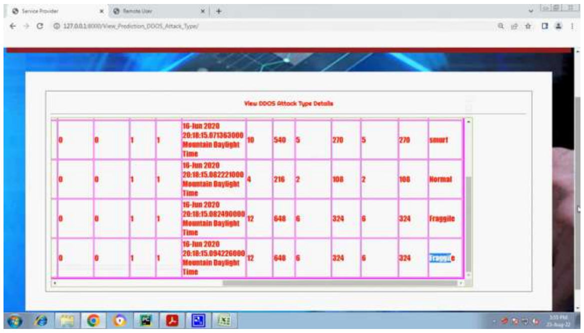
View DDOS Attack Type Details

C1912850431	udp	192.168.1.1	192.168.23.2	2419	8000	1	0	0	0	1	1
C1026483832	tcp	192.168.1.1	192.168.23.2	2426	8000	1	0	0	0	1	1
C1574873161	udp	192.168.1.1	192.168.23.2	2667	8000	1	0	0	0	1	1
C973936431	tcp	192.168.1.1	192.168.23.2	2686	8000	1	0	0	0	1	1



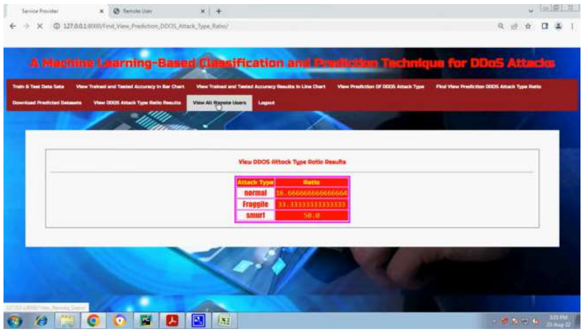
Enter RID		Enter Protocol	
Enter ip_src		Enter ip_dst	
Enter pro_srcport		Enter pro_dstport	
Enter flags_ack	--Select--	Enter ip_flags_mf	--Select--
Enter ip_flags_df	--Select--	Enter ip_flags_rb	--Select--
Enter pro_seq	--Select--	Enter pro_ack	--Select--
Enter frame_time		Enter Packets	
Enter Bytes1		Enter Tx_Packets	
Enter Tx_Bytes		Enter Rx_Packets	
Enter Rx_Bytes			

Predicted DDOS Attack Type: Fraggile



View DDOS Attack Type Details

0	0	1	1	16-Jun 2020 20:18:15.071363000 Mountain Daylight Time	10	540	5	270	5	270	smurf
0	0	1	1	16-Jun 2020 20:18:15.082221000 Mountain Daylight Time	4	216	2	108	2	108	Normal
0	0	1	1	16-Jun 2020 20:18:15.082490000 Mountain Daylight Time	12	648	6	324	6	324	Fraggile
0	0	1	1	16-Jun 2020 20:18:15.094226000 Mountain Daylight Time	12	648	6	324	6	324	Fraggile


Service Provider
Remote User
127.0.0.1:8000/Find_View_Prediction_DDOS_Attack_Type_Ratio/
A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks
Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results in Line Chart
View Prediction Of DDOS Attack Type
Find View Prediction DDOS Attack Type Ratio
Download Predicted Datasets
View DDOS Attack Type Ratio Results
View All Remote Users
Logout

View DDOS Attack Type Ratio Results

Attack Type	Ratio
normal	16.666666666666664
Fraggile	33.33333333333333
smurf	50.0

V.CONCLUSION: In this research, we suggested a comprehensive, methodical methodology to DDOS assault detection. The UNSW-nb15 dataset, which includes details on the DDOS attacks, was first chosen from the GitHub source. The Australian Centre for Cyber Security (ACCS) supplied this dataset. Data wrangling was then done using a notebook that included Python and Jupyter. Second, the dataset was separated into two classes: the independent class and the dependent class. For the algorithm, we also normalized the dataset. Following the normalization of the data, we used the suggested supervised machine learning method. The supervised method produced classification and prediction results for the model. Next, we applied the classification algorithms XG Boost and Random Forest. We found that the Random Forest Precision (PR) and Recall (RE) are both roughly 89% accurate in the initial classification. Additionally, we observed that the suggested model had an average Accuracy (AC) of about 89%, which is very fantastic and sufficient.

REFERENCES: [1] N. Martins, J. M. Cruz, T. Cruz, and P. H. Abreu, ``Adversarial machine learning applied to intrusion and malware scenarios: A systematic review," IEEE Access, vol. 8, pp. 35403_35419, 2020.

[2] G. Karatas, O. Demir, and O. K. Sahingoz, ``Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset," IEEE Access, vol. 8, pp. 32150_32162, 2020.

[3] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, ``BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset," IEEE Access, vol. 8, pp. 29575_29585, 2020.

[4] H. Jiang, Z. He, G. Ye, and H. Zhang, ``Network intrusion detection based on PSO-xgboost model," IEEE Access, vol. 8, pp. 58392_58401, 2020.

[5] A. Nagaraja, U. Boregowda, K. Khatatneh, R. Vangipuram, R. Nuvvusetty, and V. S. Kiran, ``Similarity based feature

transformation for network anomaly detection," IEEE Access, vol. 8, pp. 39184_39196, 2020.

[6] L. D'hooge, T. Wauters, B. Volckaert, and F. De Turck, ``Classi_cation hardness for supervised learners on 20 years of intrusion detection data," IEEE Access, vol. 7, pp. 167455_167469, 2019.

[7] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, ``An adaptive ensemble machine learning model for intrusion detection," IEEE Access, vol. 7, pp. 82512_82521, 2019.

[8] Y. Yang, K. Zheng, B. Wu, Y. Yang, and X. Wang, ``Network intrusion detection based on supervised adversarial variational auto-encoder with regularization," IEEE Access, vol. 8, pp. 42169_42184, 2020.

[9] C. Liu, Y. Liu, Y. Yan, and J. Wang, ``An intrusion detection model with hierarchical attention mechanism," IEEE Access, vol. 8, pp. 67542_67554, 2020.

[10] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, ``Toward a lightweight intrusion detection system for the Internet of Things," IEEE Access, vol. 7, pp. 42450_42471, 2019.