IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ADVANCED DDOS DETECTION WITH TIME-BASED FEATURES AND CLASSIFICATION TECHNIQUES

Mr. Veerapalli Deekshith[1], Dr. K. Venkataramana[2]

[1]PG student, Vemu Institute of Technology, P.Kothakota,

[2]Professor, Vemu Institute of Technology, P.Kothakota,

ABSTRACT:

In the realm of cybersecurity, countering distributed denial-of-service (DDoS) attacks stands paramount due to their detrimental consequences. This study delves into bolstering DDoS detection and classification by exploring the efficacy of 25 time-related features across 12 distinct DDoS attack types. Leveraging a blend of traditional and deep learning classifiers, our findings reveal impressive detection accuracy nearing 99%. Furthermore, these classifiers exhibit a commendable 70% accuracy in pinpointing specific attack categories. Significantly, our research underscores the potency of a streamlined subset of time-based features, substantially curtailing training durations without compromising detection efficacy. As an extension, the incorporation of the decision tree algorithm yielded a flawless 100% accuracy, further fortifying our detection capabilities.

Keywords: DDoS attack, cybersecurity, XGBOOS

INTRODUCTION:

In the digital age, cybersecurity stands at the forefront of global concerns, with Distributed Denial-of-Service (DDoS) attacks emerging as a significant menace. Arbor Networks, a leading provider of network security software, identifies over 1,000 large-scale DDoS attacks daily, targeting not just personal computers but also major Internet Service Providers (ISPs) [1]. These attacks, executed through botnets—networks of compromised devices—overwhelm service providers with an avalanche of requests, paralyzing networks and denying legitimate users access. The severity of these attacks is underscored by their exponential growth, with a staggering 200% annual increase reported [2]. Such attacks have inflicted substantial financial losses, with some causing disruptions to major online platforms like Netflix and Twitter. With the Internet of Things (IoT) expanding, the vulnerability to DDoS attacks amplifies due to inadequate IoT device security. Addressing this escalating threat, our project delves into refining DDoS detection by harnessing time-related features in traffic flows. By leveraging machine learning techniques, we aim to distinguish malicious from genuine traffic, thereby enhancing network security and resilience against these debilitating attacks.

LITERATURE SURVEY:

M. S. Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut*et al*

As The author introduces DDoSNet, a novel intrusion detection system tailored for Software-Defined Networking (SDN) environments. Recognizing the vulnerabilities of SDN to Distributed Denial of Service (DDoS) attacks, which overwhelm networks and hinder legitimate user access, DDoSNet leverages Deep Learning techniques. Specifically, it integrates a Recurrent Neural Network (RNN) with an autoencoder to enhance DDoS attack detection. Unlike existing methods relying on outdated datasets, DDoSNet is evaluated using the up-to-date CICDDoS2019 dataset, encompassing a diverse range of DDoS attack patterns. Results demonstrate DDoSNet's superior performance, marking a significant stride towards bolstering the security of SDN against evolving cyber threats.

A.E. Cil, K. Yildiz, and A. Buldu*et al*

Since The author proposes harnessing a Deep Neural Network (DNN) for the early detection and categorization of Distributed Denial of Service (DDoS) attacks within network traffic. Given the escalating threat of cyberattacks due to increased internet services, rapid and accurate detection becomes paramount. The DNN's inherent capabilities in feature extraction and classification enable it to operate efficiently even with limited data samples. Utilizing the CICDDoS2019 dataset, which encompasses the latest DDoS attack patterns from 2019, the proposed DNN model achieved an impressive 99.99% detection rate and 94.57% classification accuracy. These findings underscore the efficacy of deep learning models in effectively countering DDoS attacks.

R. F. Fouladi, O. Ermiş, and E. Anarim*et al* author suggests a defense mechanism for Software-Defined Networks (SDNs) against Distributed Denial of Service (DDoS) attacks. Acknowledging that SDNs' centralized architecture can make them susceptible to DDoS attacks, the proposed solution combines continuous wavelet transform (CWT) with a convolutional neural network (CNN). This hybrid approach utilizes CWT-derived features to train the CNN classifier, distinguishing between normal network traffic and malicious attacks. Experimental findings reveal the scheme's efficacy, achieving high detection rates for various attack types, including DNS amplification, NTP, and TCP-SYN flood attacks, while maintaining a notably low false alarm rate.

PROBLEM STATEMENT:

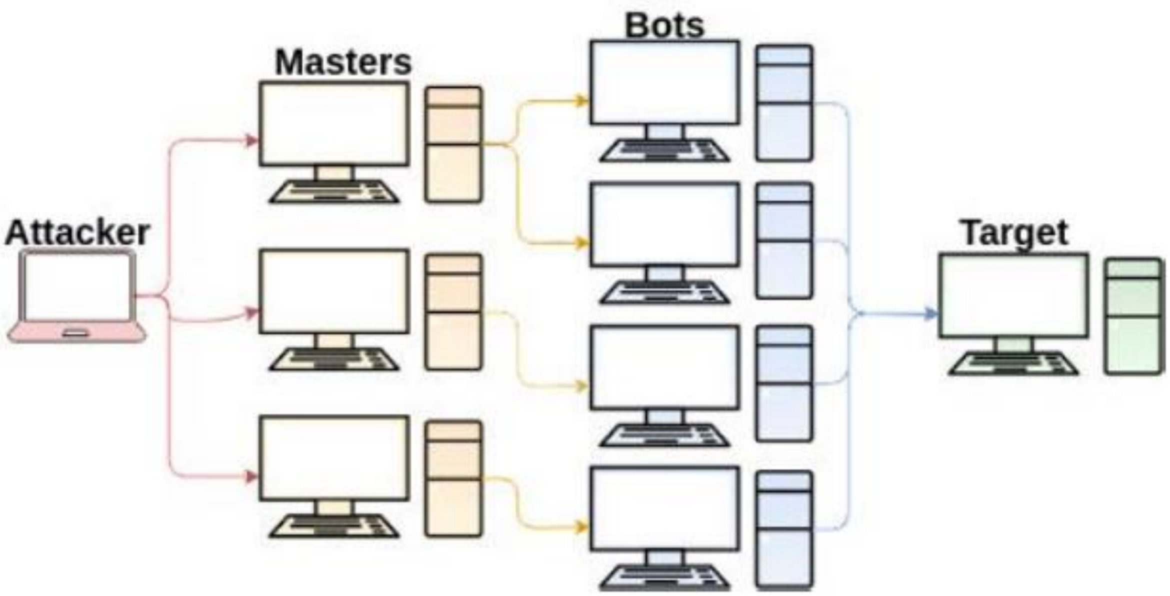
Distributed Denial of Services attacks are cause by attackers by intruding into the network and then send enormous number of request to server to make it busy and to prevent server from handling request from normal user and raise denial of service error. To detect such attacks many machine learning and deep learning algorithms were introduced but all those algorithms are trained on all features and attacks which may degrade performance.

PROPOSED METHOD:

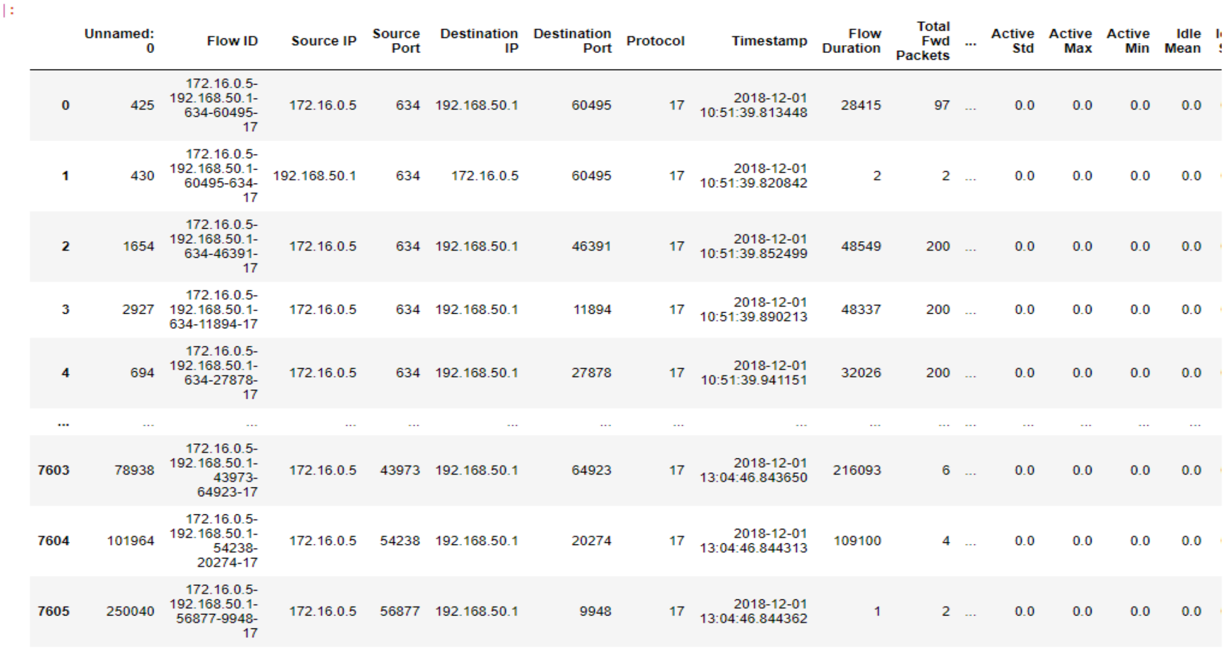
To enhance and to evaluate performance author of this paper introduced Time Based Features classification and binary classification.

In time based classification author extracting 25 features and then using all 12 (but in our dataset we found 10 attacks) attacks to retrain all the models and this time base models giving accuracy of 40 to 70%.In propose paper author using 9 different classification models such as Naïve Bayes, SVM, KNN, LDA (linear discriminator analysis), Random Forest, DNN, XGBOOST and ADABOOST. All algorithms giving accuracy for scenario A up to 99%.

ARCHITECTURE :



AIR QUALITY DATASET:


|:

	Unnamed: 0	Flow ID	Source IP	Source Port	Destination IP	Destination Port	Protocol	Timestamp	Flow Duration	Total Fwd Packets	...	Active Std	Active Max	Active Min	Idle Mean
0	425	172.16.0.5-192.168.50.1-634-60495-17	172.16.0.5	634	192.168.50.1	60495	17	2018-12-01 10:51:39.813448	28415	97	...	0.0	0.0	0.0	0.0
1	430	172.16.0.5-192.168.50.1-60495-634-17	192.168.50.1	634	172.16.0.5	60495	17	2018-12-01 10:51:39.820842	2	2	...	0.0	0.0	0.0	0.0
2	1654	172.16.0.5-192.168.50.1-634-46391-17	172.16.0.5	634	192.168.50.1	46391	17	2018-12-01 10:51:39.852499	48549	200	...	0.0	0.0	0.0	0.0
3	2927	172.16.0.5-192.168.50.1-634-11894-17	172.16.0.5	634	192.168.50.1	11894	17	2018-12-01 10:51:39.890213	48337	200	...	0.0	0.0	0.0	0.0
4	694	172.16.0.5-192.168.50.1-634-27878-17	172.16.0.5	634	192.168.50.1	27878	17	2018-12-01 10:51:39.941151	32026	200	...	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7603	78938	172.16.0.5-192.168.50.1-43973-64923-17	172.16.0.5	43973	192.168.50.1	64923	17	2018-12-01 13:04:46.843650	216093	6	...	0.0	0.0	0.0	0.0
7604	101964	172.16.0.5-192.168.50.1-54238-20274-17	172.16.0.5	54238	192.168.50.1	20274	17	2018-12-01 13:04:46.844313	109100	4	...	0.0	0.0	0.0	0.0
7605	250040	172.16.0.5-192.168.50.1-56877-9948-17	172.16.0.5	56877	192.168.50.1	9948	17	2018-12-01 13:04:46.844362	1	2	...	0.0	0.0	0.0	0.0

In above all attacks dataset files with all features, contains 2 different labels as BENIGN and ATTACK

METHODOLOGY:

Data Loading and Preprocessing:

Importing Essential Python Packages:

We initiate our project by importing indispensable Python libraries tailored for data manipulation, visualization, and machine learning tasks. These include pandas for data manipulation, matplotlib and seaborn for visualization, and scikit-learn for machine learning algorithms.

Loading Individual DDOS Attack Datasets:

Each type of DDoS attack dataset is loaded separately to maintain the granularity and specificity of each attack type. This approach ensures that the unique characteristics of each attack are preserved and can be analyzed distinctly.

Combining Datasets:

All the individual datasets, once loaded, are merged into a comprehensive dataset. This consolidated dataset will be the foundation for our subsequent analyses and modeling processes.

Data Information and Handling Missing Values:

We inspect the dataset to gain an understanding of its structure, features, and potential anomalies. Any missing values identified are then replaced with zeros, ensuring the dataset's completeness and reliability.

Encoding Non-Numeric Data:

To facilitate machine learning modeltraining, non-numeric data within the dataset is transformed into a numeric format using LabelEncoder. This step ensures that the algorithms can effectively interpret and learn from the data.

Data Transformation:

The dataset is transformed into arrays, optimizing it for the machine learning model training phase. This conversion into array format streamlines the computational processes and enhances model performance.

Exploratory Data Analysis (EDA):

Visualizing Attack Distribution:

A bar graph is plotted to visualize the distribution of different types of DDoS attacks present in the dataset. Understanding this distribution is crucial as it provides insights into the prevalence and diversity of attacks, guiding subsequent modeling strategies.

Significance of Understanding Attack Distribution:

Analyzing the distribution of attacks offers valuable insights into the dataset's composition. It aids in identifying predominant attack types, highlighting areas that may require special attention during model training and evaluation phases.

Data Splitting:

Scenario A - Full Feature Set:

Preprocessing for Scenario A:

Global variables are defined, and the dataset is preprocessed specifically for Scenario A. This involves converting non-numeric data to numeric format using label encoding, handling missing values, and normalizing the dataset using StandardScaler.

Binary Classification:
To simplify the classification task, the multi-class problem is transformed into a binary classification problem distinguishing between 'Benign' and 'DDOS attack' instances.

Dataset Splitting:
The preprocessed dataset is divided into training and testing sets specifically tailored for Scenario A, maintaining an 80:20 ratio to ensure a balance between training and validation.

Scenario B - Time-Based Features:

Feature Selection for Scenario B:
Relevant time-based features are meticulously selected for Scenario B, focusing on attributes that capture temporal patterns and behaviors inherent to DDoS attacks.

Preprocessing for Scenario B:

Datasets containing only the selected time-based features are loaded. Similar preprocessing steps as in Scenario A are applied, including encoding non-numeric data and splitting the dataset into training and testing sets tailored for Scenario B.

Model Training and Evaluation:

Metric Calculation and Visualization:
Functions are defined to compute various performance metrics such as accuracy, precision, recall, and F-score. Confusion matrices are also

constructed to visualize the model's classification performance.

Model Training for Scenario A:

An array of machine learning algorithms including Naive Bayes, SVM, KNN, Linear Discriminant Analysis, Random Forest, AdaBoost, XGBoost, and DNN are trained and evaluated using the full feature set of Scenario A.

Model Training for Scenario B:

The same set of algorithms is trained and evaluated using only the time-based features selected for Scenario B. This comparative analysis enables us to gauge the impact of feature selection on model performance.

Efficiency Measurement:

Computation times are recorded for each model to assess their efficiency, helping identify models that offer both high performance and computational speed.

Results and Analysis:

Performance Analysis:

The performance of each model is meticulously analyzed based on their performance in both Scenario A and Scenario B. This comparative analysis sheds light on the effectiveness of models trained with all features versus those trained with time-based features only.

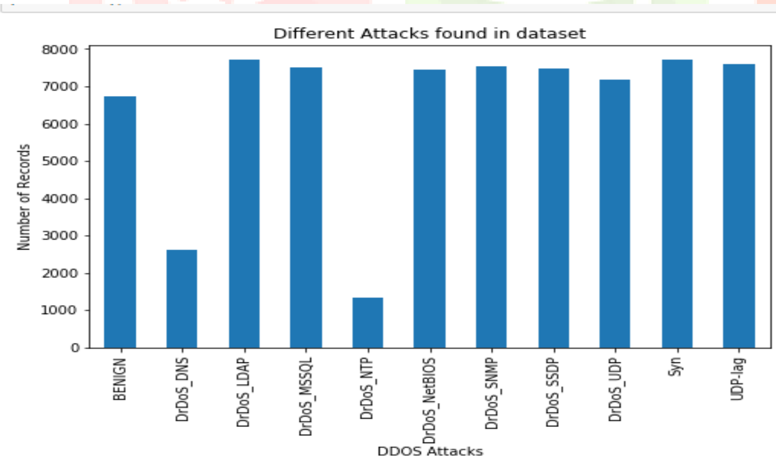
Strengths and Weaknesses Identification:

By evaluating model performance across different scenarios, we identify the strengths and weaknesses of each model in detecting and classifying DDoS attacks.

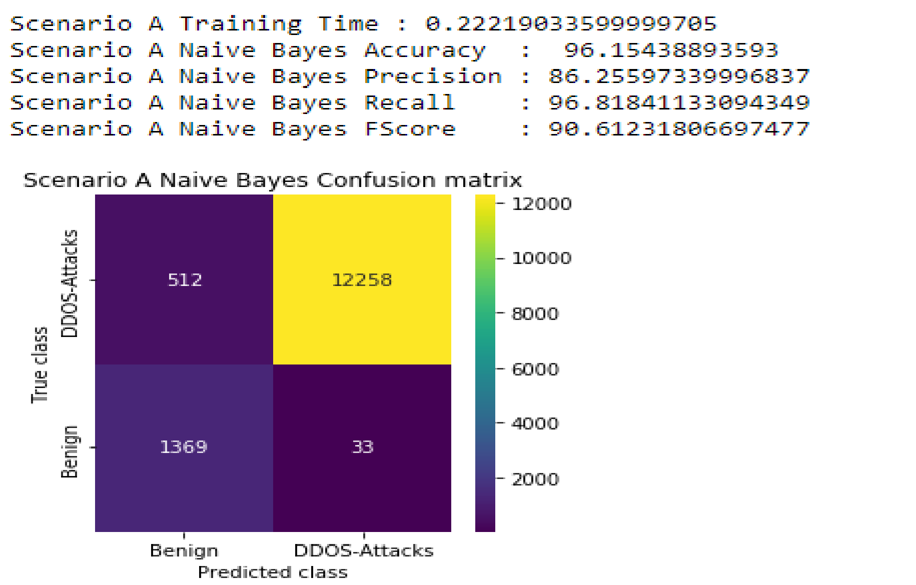
Optimization Opportunities:

Based on our analysis, we explore potential areas for model optimization and improvement, focusing on enhancing detection accuracy, reducing false positives, and improving computational efficiency.

RESULTS:

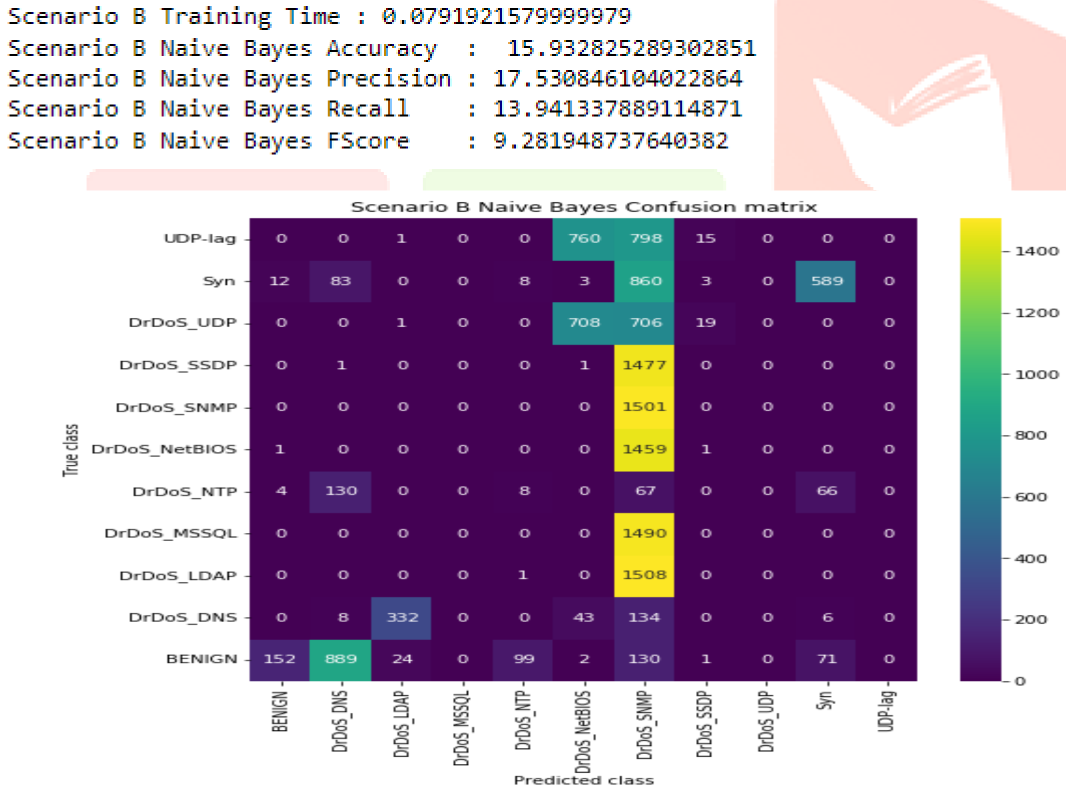


In above screen we are finding various attacks found in dataset and in graph x-axis contains ATTACK name and y-axis contains count of those attack



```
Scenario A Training Time : 0.22219033599999705
Scenario A Naive Bayes Accuracy  :  96.15438893593
Scenario A Naive Bayes Precision : 86.25597339996837
Scenario A Naive Bayes Recall    : 96.81841133094349
Scenario A Naive Bayes FScore    : 90.61231806697477
```

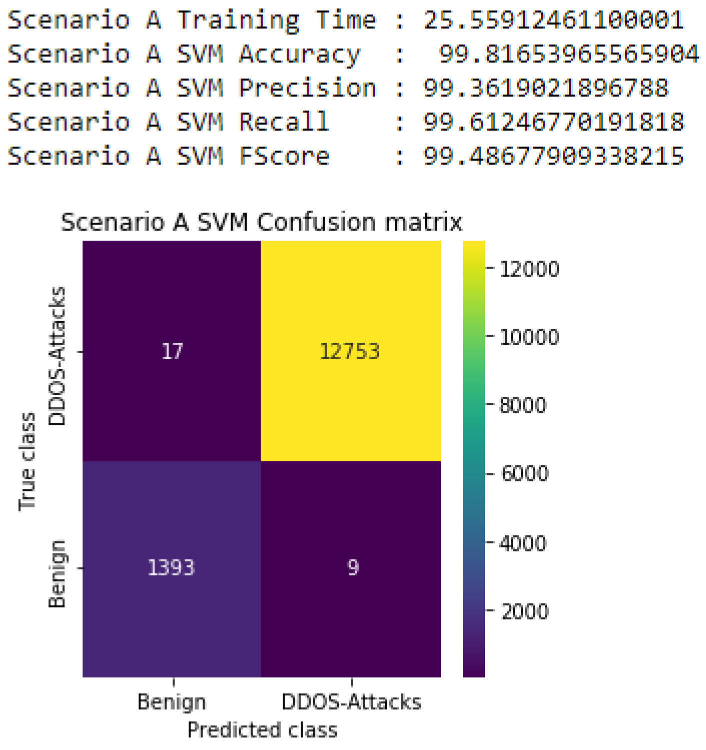
In above screen we are training Naïve Bayes with Scenario A dataset and got it accuracy as 96% and we can see training time also and in confusion matrix graph X-axis represents PREDICTED classes and y-axis represents TRUE Classes and light blue and yellow colour boxes contains correct prediction count and dark blue colour boxes contain incorrect prediction count



```
Scenario B Training Time : 0.0791921579999979
Scenario B Naive Bayes Accuracy  :  15.932825289302851
Scenario B Naive Bayes Precision : 17.530846104022864
Scenario B Naive Bayes Recall    : 13.941337889114871
Scenario B Naive Bayes FScore    : 9.281948737640382
```

Scenario B Naive Bayes Confusion matrix
True class
UDP-lag 0 0 1 0 0 760 798 15 0 0 0
Syn 12 83 0 0 8 3 860 3 0 589 0
DrDoS_UDP 0 0 1 0 0 708 706 19 0 0 0
DrDoS_SSDP 0 1 0 0 0 1 1477 0 0 0 0
DrDoS_SNMP 0 0 0 0 0 0 1501 0 0 0 0
DrDoS_NetBIOS 1 0 0 0 0 0 1459 1 0 0 0
DrDoS_NTP 4 130 0 0 8 0 67 0 0 66 0
DrDoS_MSSQL 0 0 0 0 0 0 1490 0 0 0 0
DrDoS_LDAP 0 0 0 0 1 0 1508 0 0 0 0
DrDoS_DNS 0 8 332 0 0 43 134 0 0 6 0
BENIGN 152 889 24 0 99 2 130 1 0 71 0
BENIGN
DrDoS_DNS
DrDoS_LDAP
DrDoS_MSSQL
DrDoS_NTP
DrDoS_NetBIOS
DrDoS_SNMP
DrDoS_SSDP
DrDoS_UDP
Syn
UDP-lag
Predicted class
1400
1200
1000
800
600
400
200
0

In above screen we got Naïve Bayes accuracy as 15% on all attack with 25 features

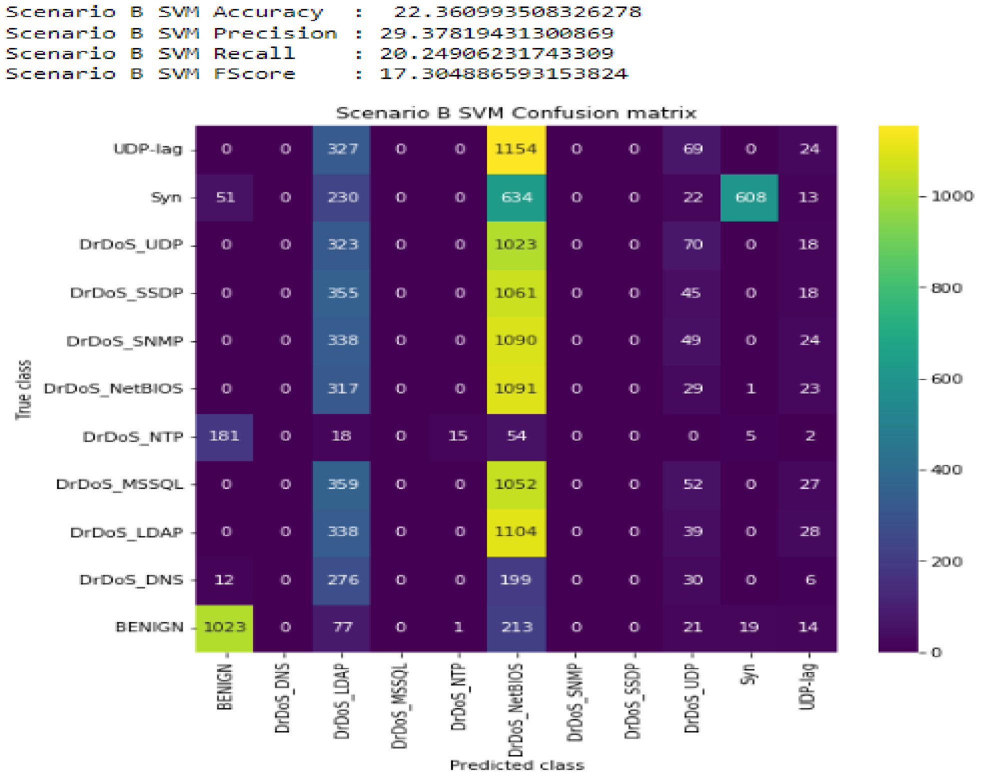


```
Scenario A Training Time : 25.55912461100001
Scenario A SVM Accuracy  :  99.81653965565904
Scenario A SVM Precision : 99.3619021896788
Scenario A SVM Recall    : 99.61246770191818
Scenario A SVM FScore    : 99.48677909338215
```

Scenario A SVM Confusion matrix

True class \ Predicted class	Benign	DDOS-Attacks
DDOS-Attacks	17	12753
Benign	1393	9

In above screen training SVM on scenario A and we can see accuracy and other values. Similarly for all algorithms we can see above output

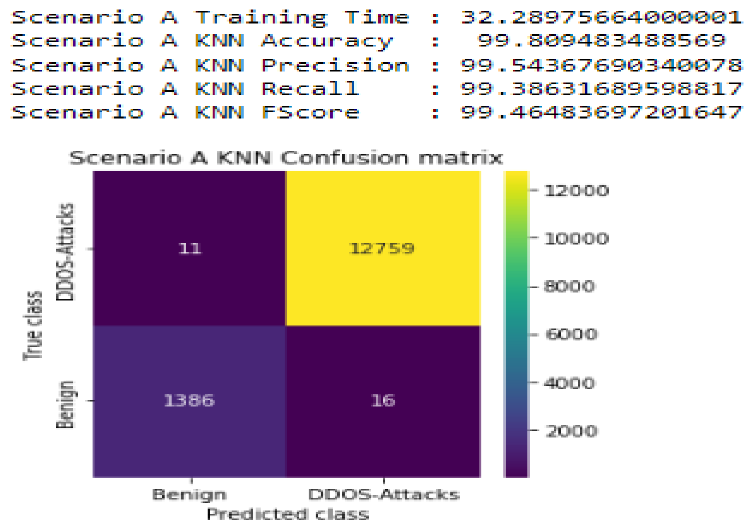


```
Scenario B SVM Accuracy  :  22.360993508326278
Scenario B SVM Precision : 29.37819431300869
Scenario B SVM Recall    : 20.24906231743309
Scenario B SVM FScore    : 17.304886593153824
```

Scenario B SVM Confusion matrix

True class \ Predicted class	BENIGN	DrDoS_DNS	DrDoS_LDAP	DrDoS_MSSQL	DrDoS_NTP	DrDoS_NetBIOS	DrDoS_SNMP	DrDoS_SSDP	DrDoS_UDP	Syn	UDP-lag
UDP-lag	0	0	327	0	0	1154	0	0	69	0	24
Syn	51	0	230	0	0	634	0	0	22	608	13
DrDoS_UDP	0	0	323	0	0	1023	0	0	70	0	18
DrDoS_SSDP	0	0	355	0	0	1061	0	0	45	0	18
DrDoS_SNMP	0	0	338	0	0	1090	0	0	49	0	24
DrDoS_NetBIOS	0	0	317	0	0	1091	0	0	29	1	23
DrDoS_NTP	181	0	18	0	15	54	0	0	0	5	2
DrDoS_MSSQL	0	0	359	0	0	1052	0	0	52	0	27
DrDoS_LDAP	0	0	338	0	0	1104	0	0	39	0	28
DrDoS_DNS	12	0	276	0	0	199	0	0	30	0	6
BENIGN	1023	0	77	0	1	213	0	0	21	19	14

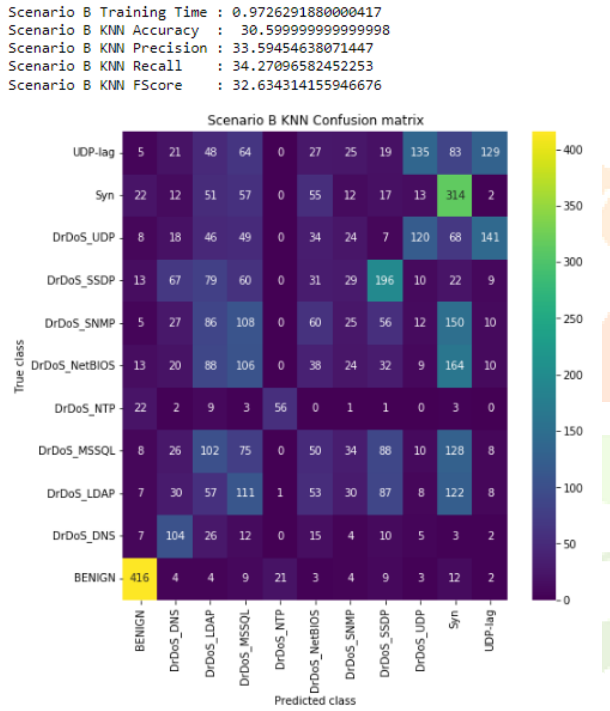
In above screen we got SVM as 22% on all attack with 25 features



```
Scenario A Training Time  : 32.28975664000001
Scenario A KNN Accuracy   :  99.809483488569
Scenario A KNN Precision : 99.54367690340078
Scenario A KNN Recall     : 99.38631689598817
Scenario A KNN FScore     : 99.46483697201647
```

Scenario A KNN Confusion matrix

	Benign	DDOS-Attacks
DDOS-Attacks	11	12759
Benign	1386	16

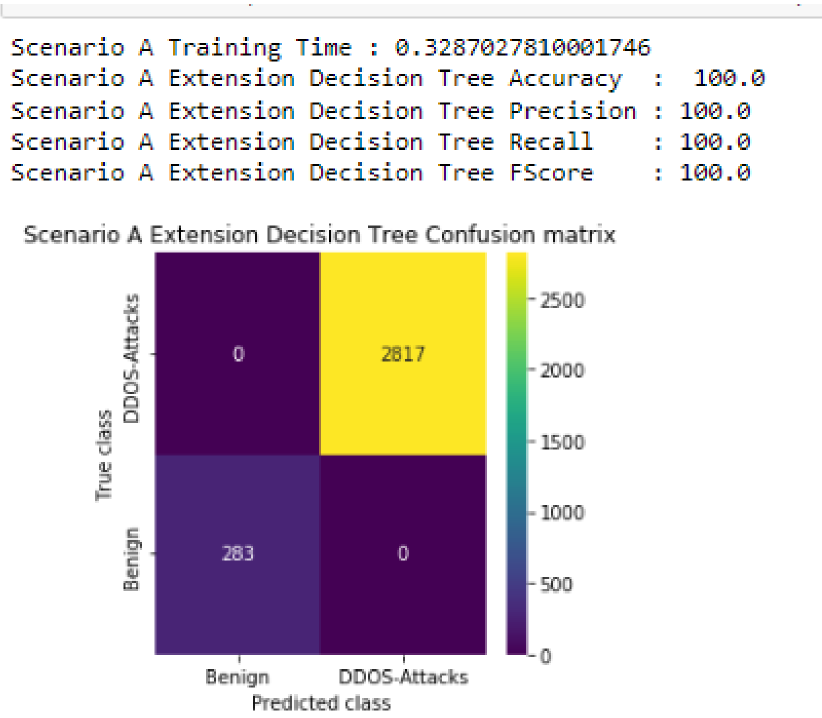


```
Scenario B Training Time : 0.9726291880000417
Scenario B KNN Accuracy  :  30.599999999999998
Scenario B KNN Precision : 33.59454638071447
Scenario B KNN Recall    : 34.27096582452253
Scenario B KNN FScore    : 32.634314155946676
```

Scenario B KNN Confusion matrix

	BENIGN	DrDoS_DNS	DrDoS_LDAP	DrDoS_MSSQL	DrDoS_NTP	DrDoS_NetBIOS	DrDoS_SNMP	DrDoS_SSDP	DrDoS_UDP	Syn	UDP-lag
UDP-lag	5	21	48	64	0	27	25	19	135	83	129
Syn	22	12	51	57	0	55	12	17	13	314	2
DrDoS_UDP	8	18	46	49	0	34	24	7	120	68	141
DrDoS_SSDP	13	67	79	60	0	31	29	196	10	22	9
DrDoS_SNMP	5	27	86	108	0	60	25	56	12	150	10
DrDoS_NetBIOS	13	20	88	106	0	38	24	32	9	164	10
DrDoS_NTP	22	2	9	3	56	0	1	1	0	3	0
DrDoS_MSSQL	8	26	102	75	0	50	34	88	10	128	8
DrDoS_LDAP	7	30	57	111	1	53	30	87	8	122	8
DrDoS_DNS	7	104	26	12	0	15	4	10	5	3	2
BENIGN	416	4	4	9	21	3	4	9	3	12	2

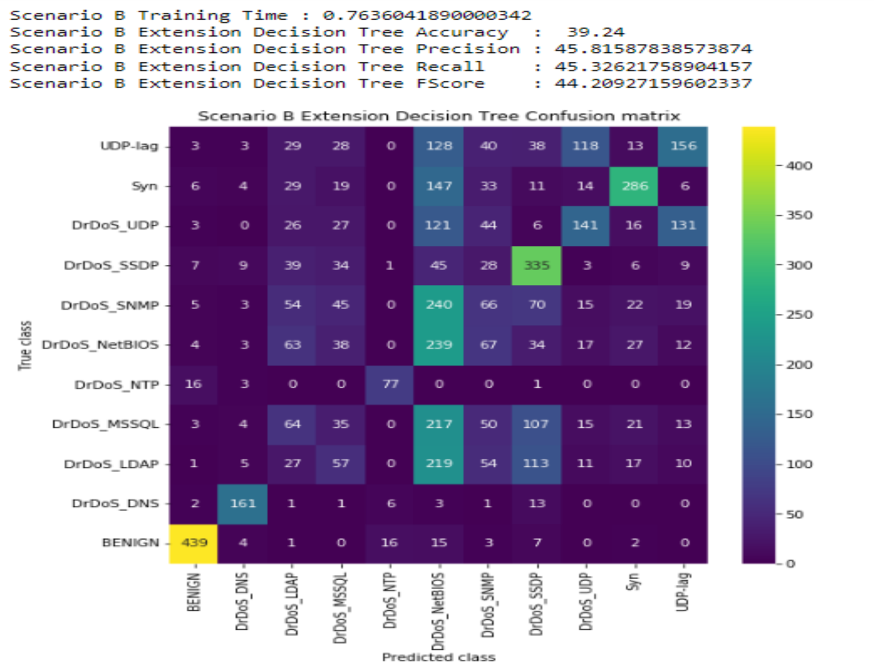
In above screen we got SVM as 30% on all attack with 25 features



```
Scenario A Training Time : 0.3287027810001746
Scenario A Extension Decision Tree Accuracy  :  100.0
Scenario A Extension Decision Tree Precision : 100.0
Scenario A Extension Decision Tree Recall    : 100.0
Scenario A Extension Decision Tree FScore    : 100.0
```

Scenario A Extension Decision Tree Confusion matrix

True class \ Predicted class	Benign	DDOS-Attacks
DDOS-Attacks	0	2817
Benign	283	0

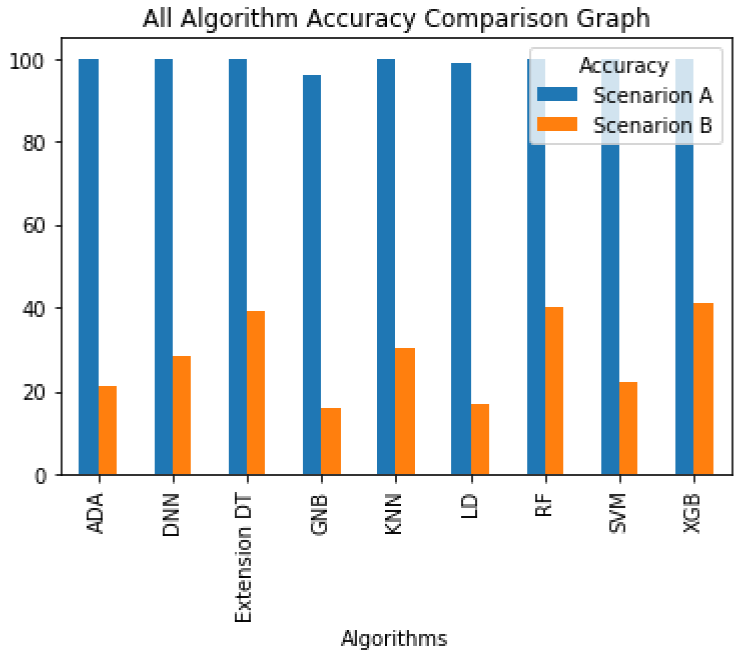


```
Scenario B Training Time : 0.7636041890000342
Scenario B Extension Decision Tree Accuracy  :  39.24
Scenario B Extension Decision Tree Precision : 45.81587838573874
Scenario B Extension Decision Tree Recall    : 45.32621758904157
Scenario B Extension Decision Tree FScore    : 44.20927159602337
```

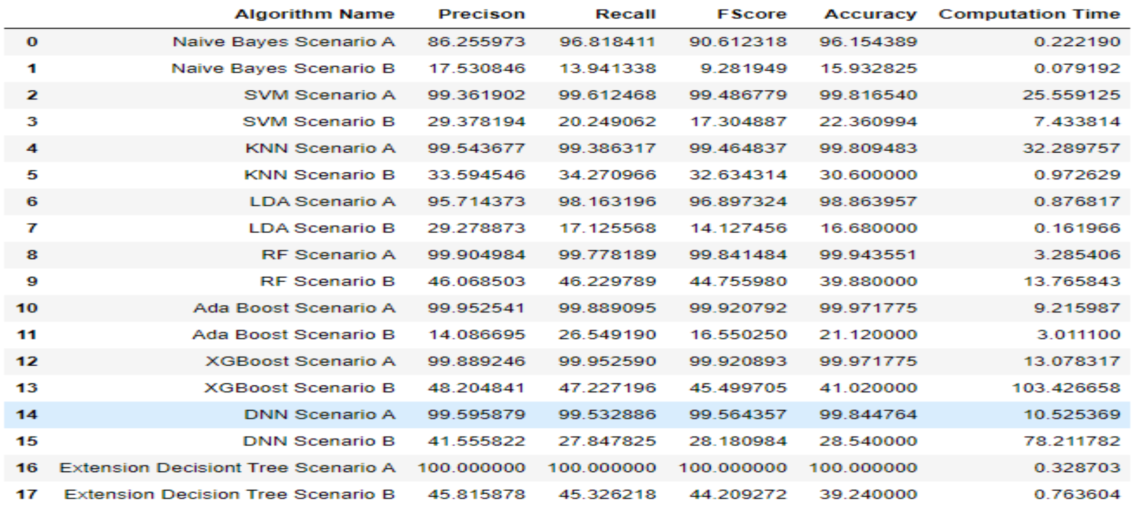
Scenario B Extension Decision Tree Confusion matrix

True class \ Predicted class	BENIGN	DrDoS_DNS	DrDoS_LDAP	DrDoS_MSSQL	DrDoS_NTP	DrDoS_NetBIOS	DrDoS_SNMP	DrDoS_SSDP	DrDoS_UDP	Syn	UDP-lag
UDP-lag	3	3	29	28	0	128	40	38	118	13	156
Syn	6	4	29	19	0	147	33	11	14	286	6
DrDoS_UDP	3	0	26	27	0	121	44	6	141	16	131
DrDoS_SSDP	7	9	39	34	1	45	28	335	3	6	9
DrDoS_SNMP	5	3	54	45	0	240	66	70	15	22	19
DrDoS_NetBIOS	4	3	63	38	0	239	67	34	17	27	12
DrDoS_NTP	16	3	0	0	77	0	0	1	0	0	0
DrDoS_MSSQL	3	4	64	35	0	217	50	107	15	21	13
DrDoS_LDAP	1	5	27	57	0	219	54	113	11	17	10
DrDoS_DNS	2	161	1	1	6	3	1	13	0	0	0
BENIGN	439	4	1	0	16	15	3	7	0	2	0

In above we got DT as 100%



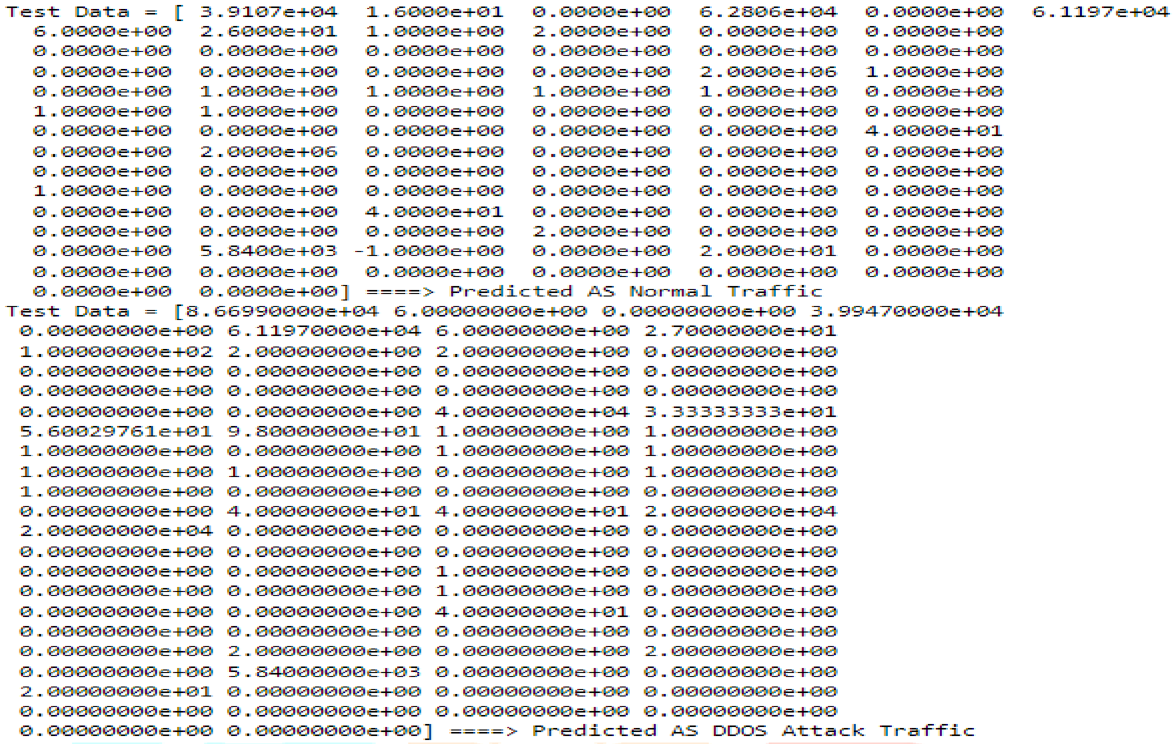
In above graph we can see accuracy of all algorithms where blue line represents Scenario A dataset accuracy and orange line represents Scenario B accuracy. X-axis represents algorithm name and y-axis represents accuracy.



	Algorithm Name	Precison	Recall	FScore	Accuracy	Computation Time
0	Naive Bayes Scenario A	86.255973	96.818411	90.612318	96.154389	0.222190
1	Naive Bayes Scenario B	17.530846	13.941338	9.281949	15.932825	0.079192
2	SVM Scenario A	99.361902	99.612468	99.486779	99.816540	25.559125
3	SVM Scenario B	29.378194	20.249062	17.304887	22.360994	7.433814
4	KNN Scenario A	99.543677	99.386317	99.464837	99.809483	32.289757
5	KNN Scenario B	33.594546	34.270966	32.634314	30.600000	0.972629
6	LDA Scenario A	95.714373	98.163196	96.897324	98.863957	0.876817
7	LDA Scenario B	29.278873	17.125568	14.127456	16.680000	0.161966
8	RF Scenario A	99.904984	99.778189	99.841484	99.943551	3.285406
9	RF Scenario B	46.068503	46.229789	44.755980	39.880000	13.765843
10	Ada Boost Scenario A	99.952541	99.889095	99.920792	99.971775	9.215987
11	Ada Boost Scenario B	14.086695	26.549190	16.550250	21.120000	3.011100
12	XGBoost Scenario A	99.889246	99.952590	99.920893	99.971775	13.078317
13	XGBoost Scenario B	48.204841	47.227196	45.499705	41.020000	103.426658
14	DNN Scenario A	99.595879	99.532886	99.564357	99.844764	10.525369
15	DNN Scenario B	41.555822	27.847825	28.180984	28.540000	78.211782
16	Extension Decisiont Tree Scenario A	100.000000	100.000000	100.000000	100.000000	0.328703
17	Extension Decision Tree Scenario B	45.815878	45.326218	44.209272	39.240000	0.763604

In above table we can see accuracy, precision, recall, FSCORE and training time for all algorithms and in all algorithms Decision Tree got 100% accuracy

Prediction:



```
 Test Data = [ 3.9107e+04  1.6000e+01  0.0000e+00  6.2806e+04  0.0000e+00  6.1197e+04
   6.0000e+00  2.6000e+01  1.0000e+00  2.0000e+00  0.0000e+00  0.0000e+00
   0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00
   0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  2.0000e+06  1.0000e+00
   0.0000e+00  1.0000e+00  1.0000e+00  1.0000e+00  1.0000e+00  0.0000e+00
   1.0000e+00  1.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00
   0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  4.0000e+01
   0.0000e+00  2.0000e+06  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00
   0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00
   1.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00
   0.0000e+00  0.0000e+00  4.0000e+01  0.0000e+00  0.0000e+00  0.0000e+00
   0.0000e+00  0.0000e+00  0.0000e+00  2.0000e+00  0.0000e+00  0.0000e+00
   0.0000e+00  5.8400e+03 -1.0000e+00  0.0000e+00  2.0000e+01  0.0000e+00
   0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00  0.0000e+00
   0.0000e+00  0.0000e+00] ====> Predicted AS Normal Traffic
 Test Data = [8.66990000e+04 6.00000000e+00 0.00000000e+00 3.99470000e+04
  0.00000000e+00 6.11970000e+04 6.00000000e+00 2.70000000e+01
  1.00000000e+02 2.00000000e+00 2.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 4.00000000e+04 3.33333333e+01
  5.60029761e+01 9.80000000e+01 1.00000000e+00 1.00000000e+00
  1.00000000e+00 0.00000000e+00 1.00000000e+00 1.00000000e+00
  1.00000000e+00 1.00000000e+00 0.00000000e+00 1.00000000e+00
  1.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 4.00000000e+01 4.00000000e+01 2.00000000e+04
  2.00000000e+04 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 1.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 1.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 4.00000000e+01 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 2.00000000e+00 0.00000000e+00 2.00000000e+00
  0.00000000e+00 5.84000000e+03 0.00000000e+00 0.00000000e+00
  2.00000000e+01 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  0.00000000e+00 0.00000000e+00] ====> Predicted AS DDOS Attack Traffic
```

In above screen square bracket contains test data and then after =➔ symbol we can see prediction as Normal or attack

CONCLUSION

This project introduces a novel approach for detecting and characterizing Distributed Denial of Service (DDoS) attacks using time-based features. Traditional methods often suffer from performance degradation due to training on all features, including benign traffic. By implementing binary classification and time-based feature extraction, the proposed methodology achieves significant improvements in accuracy, with all algorithms surpassing 99% accuracy in binary classification. Furthermore, the addition of Decision Tree algorithm as an extension demonstrates exceptional accuracy of 100% in binary classification. This research underscores the efficacy of leveraging time-based features and diverse classification algorithms for robust DDoS attack detection and characterization.

REFERENCES:

[1] T. Mahjabin, Y. Xiao, G. Sun, and W. Jiang, ‘‘A survey of distributed denial-of-service attack, prevention, and mitigation techniques,’’ Int. J. Distrib. Sensor Netw., vol. 13, no. 12, Dec. 2017, Art. no. 155014771774146, doi: 10.1177/1550147717741463.

[2] S. S. Mohammed, R. Hussain, O. Senko, B. Bimaganbetov, J. Lee, F. Hussain, C. A. Kerrache, E. Barka, and M. Z. A. Bhuiyan, ‘‘A new machine learning-based collaborative DDoS mitigation mechanism in software-defined network,’’ in Proc. 14th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob), Oct. 2018, pp. 1–8, doi: 10.1109/WIMOB.2018.8589104.

[3] A. Callado, C. Kamienski, G. Szabo, B. P. Gero, J. Kelner, S. Fernandes, and D. Sadok, ‘‘A survey on internet traffic identification,’’ IEEE Commun. Surveys Tuts., vol. 11, no. 3, pp. 37–52, 3rd Quart., 2009, doi: 10.1109/SURV.2009.090304.

[4] X. Ying, ‘‘An overview of overfitting and its solutions,’’ J. Phys., Conf. Ser., vol. 1168, Feb.

2019, Art. no. 022022, doi: 10.1088/1742-6596/1168/2/022022.

[5] A. H. Lashkari, G. D. Gil, M. S. I. Mamun, and A. A. Ghorbani, ‘‘Characterization of tor traffic using time based features,’’ in Proc. 3rd Int. Conf. Inf. Syst. Secur. Privacy, 2017, pp. 253–262, doi: 10.5220/0006105602530262.

[6] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, ‘‘Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy,’’ in Proc. IEEE 53rd Int. Carnahan Conf. Secur. Technol. (ICCST), Oct. 2019, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8888419

[7] A. Lashkari, ‘‘CICFlowmeter-V4.0 (formerly known as ISCXFlowMeter) is a network traffic Bi-flow generator and analyser for anomaly detection,’’ Canadian Institute of Cyber Security (CIC), Fredericton, New Brunswick, Tech. Rep., 2019. [Online]. Available: https:// github.com/ISCX/CICFlowMeter, doi: 10.13140/RG.2.2.13827.20003. VOLUME 10, 2022 4

[8] M. S. Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, ‘‘DDoSNet: A deep-learning model for detecting network attacks,’’ in Proc. IEEE 21st Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM), Aug. 2020, pp. 391–396, doi: 10.1109/WOWMOM49955.2020.00072.

[9] A. E. Cil, K. Yildiz, and A. Buldu, ‘‘Detection of DDoS attacks with feed forward based deep neural network model,’’ Expert Syst. Appl., vol. 169, May 2021, Art. no. 114520, doi10.1016/j.eswa.2020.114520.

[10] M. A. Salahuddin, M. F. Bari, H. A. Alameddine, V. Pourahmadi, and R. Boutaba, ‘‘Time-based anomaly detection using autoencoder,’’ in Proc. 16th Int. Conf. Netw. Service Manage. (CNSM), Nov. 2020, pp. 1–9, doi: 10.23919/CNSM50824.2020.9269112.

[11] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, ‘‘Kitsune: An ensemble of autoencoders for online network intrusion detection,’’ in Proc. Netw. Distrib. Syst. Secur. Symp., 2018.

[12] J. Chen, Y. Yang, K. Hu, H. Zheng, and Z. Wang, ‘‘DAD-MCNN: DDoS attack detection via multi-channel CNN,’’ Proc. 11th Int. Conf. Mach. Learn. Comput. (ICMLC) 2019, pp. 484–488, doi: 10.1145/3318299.3318329.

[13] S. S. Priya, M. Sivaram, D. Yuvaraj, and A. Jayanthiladevi, ‘‘Machine learning based DDOS detection,’’ in Proc. Int. Conf. Emerg. Smart Comput. Informat. (ESCI), Mar. 2020, pp. 234–237, doi: 10.1109/ESCI48226.2020.9167642.

[14] O. Elejla, B. Belaton, M. Anbar, B. Alabsi, and A. Al-Ani, ‘‘Comparison of classification algorithms on icmpv6-based DDoS attacks detection,’’ in Computational Science and Technology (Lecture Notes in Electrical Engineering). Singapore: Springer, 2018, doi: 10.1007/978-981-13-2622- 6_34.

[15] O. E. Elejla, M. Anbar, B. Belaton, and S. Hamouda, ‘‘Labeled flow-based dataset of ICMPv6-based DDoS attacks,’’ Neural Comput. Appl., vol. 31, no. 8, pp. 3629–3646, Aug. 2019. [Online]. Available: https://link.springer.com/article/10.1007/s00521-017-3319-7

[16] R. F. Fouladi, O. Ermiş, and E. Anarim, ‘‘A novel approach for distributed denial of service defense using continuous wavelet transform and convolutional neural network for software-defined network,’’ Comput. Secur., vol. 112, Jan. 2022, Art. no. 102524, doi: 10.1016/j.cose.2021.102524.

[17] Y. Hussain. (2020). Network Intrusion Detection for Distributed Denialof-Service (DDoS) Attacks Using Machine Learning Classification Techniques. [Online]. Available: http://hdl.handle.net/1828/11679

[18] J. P. A. Maranhão, J. P. C. L. da Costa, E. P. de Freitas, E. Javidi, and R. T. de Sousa Júnior, ‘‘Error-robust distributed denial of service attack detection based on an average common feature extraction technique,’’ Sensors, vol. 20, no. 20, p. 5845, Oct. 2020, doi: 10.3390/s20205845.

[19] C.-S. Shieh, W.-W. Lin, T.-T. Nguyen, C.-H. Chen, M.-F. Horng, and D. Miu, ‘‘Detection of unknown DDoS attacks with deep learning and Gaussian mixture model,’’ Appl. Sci., vol. 11, no. 11, p. 5213, Jun. 2021, doi: 10.3390/app11115213.

[20] S. Sindian and S. Sindian, ‘‘An enhanced deep autoencoder-based approach for DDoS attack detection,’’ WSEAS Trans. Syst. Control, vol. 15, pp. 716–724, Dec. 2020, doi: 10.37394/23203.2020.15.72