2022 International Conference on Advances in Computing, Communication and Materials (ICACCM) | 978-1-6654-7439-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICACCM56405.2022.10009400

An Intrusion Detection Model for CICIDS-2017 Dataset Using Machine Learning Algorithms

Shailesh Singh Panwar[1]
Dept. of Computer Science and Engg.
H.N.B. Garhwal University Srinagar Garhwal, Uttarakhand, India
Shaileshpanwar23@gmail.com[1]

Y. P. Raiwani[2]
Dept. of Computer Science and Engg.
H.N.B. Garhwal University Srinagar Garhwal, Uttarakhand, India
yp_raiwani@yahoo.com[2]

Lokesh Singh Panwar[3]
Dept. of Electronics and Comm. Engg.
Graphic Era University, Dehradun, Uttarakhand, India
lokesh31j@gmail.com[3]

Abstract **-** Due to excessive use of the internet, keeping the network secure and transferring data securely across networks has become a difficult task. To recognize distinct types of network (internet) attacks, need Machine learning strategies, to develop various types of the intrusion detection system. The main reason behind using different types of intrusion detection strategies is that the attackers constantly receiving information about the network traffic. Therefore, the intrusion detection system is used to easily identify the attacks and to thwart those attacks.

In this paper, we have used the CICIDS-2017 dataset, which is a labeled dataset, for analyzing the result of eight different supervised classification techniques (GaussianNB (GNB), BernoulliNB (BNB), Decision Tree, KNN, Logistic Regression, SVM, Random Forest and SGD). The dataset is divided into five different days; each has different types of class attacks (DoS, PortScan, Botnet, Web Attacks, Infiltration, Heartbleed and DDoS). We have proposed a model with three stages. The first stage data-preprocessing, which includes several steps such as extracting independent and dependent variable, splitting dataset, finding missing value, feature scaling on the dataset and encoding categorical data, then in feature selection stage, finding out unique features using RFE (Recursive feature elimination) technique. Finally, test the result of eight different supervised classification techniques using cross-validation on the CICIDS-2017 dataset.

Keywords: Data Preprocessing, Feature Selection, Intrusion Detection System, CICIDS-2017, Machine Learning Algorithms.

I. INTRODUCTION

Currently, the use of computer networks and computer applications to share, store and capture data is constantly increasing [1]. Due to the different types of malicious attacks and the large size of the computer network, preserving both information and communication has become a challenging task for the researcher. So, in many system firewalls are used to prevent malicious attacks. IDS play an important role to increase the level of network security, which is accepted as a second line of protection to remove the threats and malicious attacks. Attackers are always looking for new ways to harm the prevention mechanism and bypass the system. Hence, IDSs becomes an important component of safety [2, 3, 4].

Intrusion Detection Systems could be a hardware device or software-based application [5, 6]. IDSs are divided into signature-based techniques and anomaly- based techniques. IDSs developers are working on various technologies and techniques to find out the intrusion detection. Out of which, machine learning is one of the most popular techniques. Techniques that can detect, determine, and predict attacks before results are called machine learning. If enough training data available and the model is sufficiently generality to detect threats, malicious attack types and novel attacks, then, can get satisfactory and novel results from the machine Learning IDSs model [7].

Machine learning instances are classified into two categories: (i) the instances are classified into two classes (Attack and Benign), called Binary Classification. (ii) If classifying instances have three and more than three classes (Benign, FTP-Patator, SSH-Patator, etc.) it is called Multiclass classification.

The remaining part of this paper is categorized as follows. In Section-II we have presented a literature review on algorithms and approaches based on the machine learning algorithms in tabulation form. In Section-III we have presented the overview of the CICIDS-2017 datasets, which is used for intrusion detection. Section IV describes feature selection and the outline of data preprocessing and it's subpart with the help of our proposed model. In Section V we have discuss in brief about the machine learning algorithms used. In Section VI we have computed the performance evaluation of machine learning algorithms used in binary and multiclass classification with data preprocessing and feature selection and without data preprocessing and feature selection. Lastly, we have concluded our work and describe the future scope in Section VII.

II. Literature Review

Intrusion Detection Systems (IDSs) is a very important safety tool against the constantly developing and growing network attacks.

Due to the need for reliable testing and effective dataset, accurate methods for anomaly-based intrusion detection and coherent representations are growing consistently.

Table 1. Related Machine Learning Algorithms for Intrusion Detection System

Author	ML Model	Work Done	Performance Metrics	Year	Dataset
Tang et al. [9]	SDN controller	For Anomaly detection used deep learning techniques in SDN environment	Accuracy, Precision, Recall, F-Measure	2016	NSL-KDD
Kim et al. [10]	RNN	Constructed an IDS with the deep learning techniques	False Alarm Rate, Detection Rate	2016	KDD Cup 1999
A. Javaid et al. [11]	Deep Learning	Developed a flexible and efficient IDSs using deep learning algorithm	Accuracy, Precision, Recall, F-Measure	2016	NSL-KDD
Yin et al. [12]	RNN	Proposed a model and compare with different machine learning algorithms	True Positive Rate, Accuracy, False Positive Rate	2017	NSL KDD
Marir et al. [13]	SVM, DBN	Developed an approach for attack detection using DBN and SVM	Precision, Recall, F- Measure	2018	KDD99, NSL-KDD, UNSWNB15
Zhou et al. [14]	RM, LR, KNN, SVM,CNN	Developed a model for online transaction using CNN	Accuracy, Precision, Recall	2018	Commercial Bank Data
Tang et al. [15]	RNN	Proposed a GRU-RNN model to IDSs for SDNS	Recall, F1 Score, Precision, Accuracy	2018	NSL-KDD
Jiang et al. [16]	RNN	Proposed an effective anomaly detection approach using neural networks	Accuracy, False Alarm Rate	2018	NSL-KDD
Aksu et al. [17]	SVM, KNN, DT	To detect DoS on CICIDS-2017 dataset	Accuracy, Precision, Recall, F-Measure	2018	CICIDS-2017
Acharya et al. [18]	SVM	Analyzed and evaluated the performance of developed approach	Detection Rate, False Alarm Rate	2018	NSL-KDD
Cavusoglu et al. [19]	NB, J48, RF	Analysed and evaluated the performance of developed approach	Accuracy of DoS	2019	NSL-KDD
Negandhi et al. [20]	Random forest	To detect different networking attack and analyzed the performance of proposed model.	Accuracy	2019	NSL-KDD
Panwar et al. [21]	NB, J48, Decision Tree	Analysed of different machine learning algorithms	Precision, Accuracy, F-Measure, Recall, Time	2019	CICIDS-2017
Dutta et al. [23]	DNN	Used hybrid anomaly detection system	Accuracy, Detection Rate, False-Positive Rate, ROC, F1-Score	2020	UNSW-NB15

The problem of attack detection is a specific categorization problem, which can be successfully resolved by machine learning. The IDS is broadly depending on the various machine learning techniques. The ability of an intrusion detection system is evaluated by the precise prediction of attacks. The literature review is shown in Table 1, in which we have summarized a survey of the different machine learning techniques used, the dataset used by the researcher and the work done is also mentioned.

III. Dataset Description

Canadian Institute for Cybersecurity has published the CICIDS-2017 dataset, in which included benign and most popular update attacks. This novel data set accommodates all 11 important and compulsory benchmark, i.e. Complete Network configuration, Heterogeneity, Anonymity, Complete Interaction, Complete Traffic, Meta Data, Available Protocols, Labeled Dataset, Feature Set, Complete Capture and Attack Diversity. It also carried out updated attacks such as DoS, PortScan, Botnet, Web Attacks, Infiltration, Heartbleed, and DDoS.

The whole data set has 2,830,743 records, which have included all attacks and benign. All files corresponding to each day accommodates 79 different attributes with the label.

In this paper, for calculating the performance of CICIDS-2017 data set, we have taken 7 files separately, out of this 4 files contains the two-class (Binary) labels i.e. PortScan, Botnet, DDoS, Infiltration and benign. Three files contain multiclass labels such as FTP-Patator, SSH-Patator and benign etc used for analysing the performance of the CICIDS-2017 dataset [8]

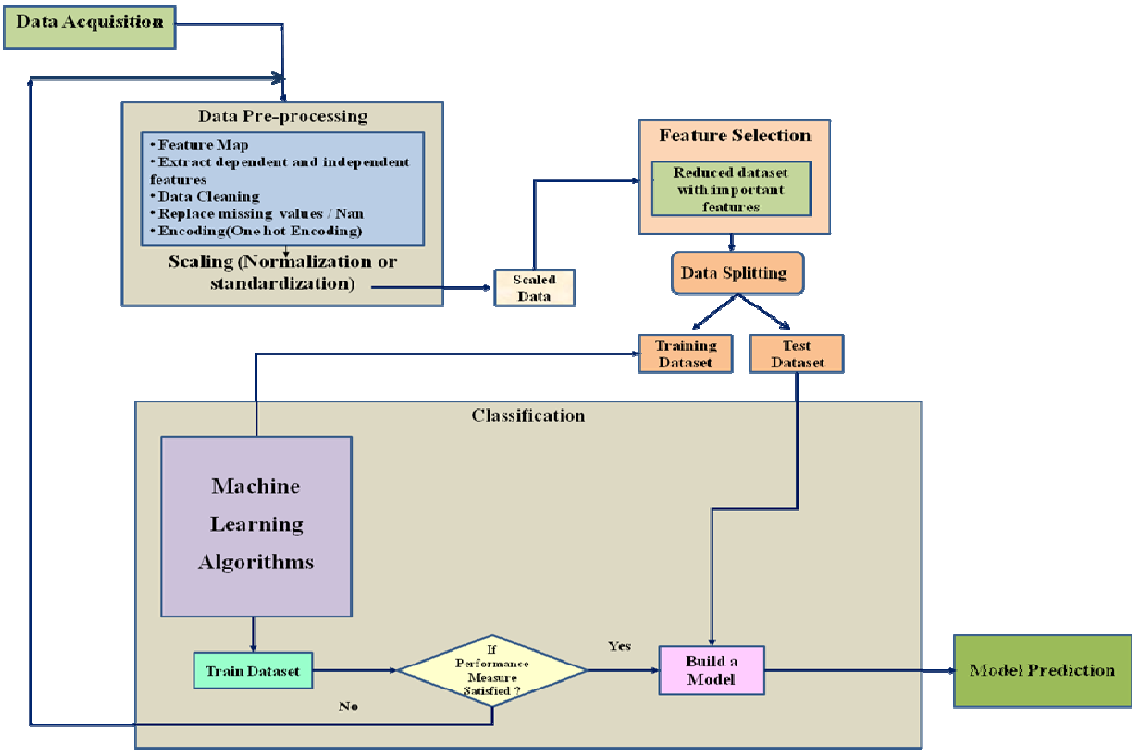

Fig 1. Proposed Model

Table 2. Total Number of Flows and Attacks par Day

Day	Total Flows	No of Attacks	Attacks Type
Monday	529918	0	Normal network activities
Tuesday	445909	7938	FTP-Patator
		5897	SSH-Patator
Wednesday	692703	5796	DoS slowloris
		5499	DoS Slowhttptest
		231073	DoS Hulk
		10293	Dos GoldenEye
		11	Heartbleed
Thursday Morning	170,366	1507	Web Attack - Brute Force
		652	Web Attack - XSS
		21	Web Attack - SQL Injection
Thursday Afternoon	288602	36	Infiltration
Friday Morning	191033	1966	Botnet
Friday Afternoon	286467	158930	PortScan
Friday Afternoon 2	225745	128027	DDoS
Total	**2830743**	**557646**	**19.70%**

IV. Experimental Setup and Result

(A) Hardware and Software used: - Our proposed model is implemented on Intel Xeon E5-2650 v4 (12 core, 2.2 GHz, 30MB L3 cache, 16 GB RAM, NVIDIA Quadro P400 2GB). We use Python (Jupyter Notebook) machine learning tool under Anaconda Navigator on Windows 10 Professional. Training and testing data is manipulated in form of Numpy arrays. Python Scikit-learn library is used to implement other classifiers for comparisons. Python programming language has been used for the development of our IDS model.

For the proposed model, using required attributes of CICIDS-2017 dataset and after going through the various steps i.e. cleaning, encoding, labeling, feature scaling and normalization, we make ensure that dataset is ready to be trained. CICIDS-2017 dataset was separated into two parts: 80% of training data and 20% of testing data.

(B) Evaluation Metrics: - To analyse the model, we have selected four parameters: - Accuracy, Recall, Precision, and F1 Score. After computing four indicators i.e. True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN); Accuracy, Recall, Precision and F1Score are evaluated.

True Positive (TP): The amount of malicious data value accurately classified into the malicious class.

False Positive) (FP): The amount of benign data value inaccurately classified into the malicious class.

False Negative (FN): The amount of malicious data value inaccurately classified into the benign class.

True Negative (TN): The amount of benign data value accurately classified into the benign class.

(i) Accuracy: It is the relationship between accurate predictions made (TP + TN) to all predictions made (TP + FP + FN + TN), namely

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

(ii) Recall: - It can be computed from the quantity of malicious attacks discovered by the network rather than the number of actual infiltration.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

(iii) Precision: - It is the relationship between the entire quantities of accurately predicted positive categories (TP) to the entire quantity of positive predictions build (TP + FP), namely

$$\text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

(iv) F1 Score (TPR):- It is the harmonic mean of the Recall and the precision.

$$\text{F1 Score} = \frac{2*\text{Recall}*\text{Precision}}{\text{Recall}+\text{Precision}} \quad (11)$$

(C) Performance Evaluation Methods: - We have implemented our model in Python. Different types of algorithms have been applied on the CICIDS-2017 dataset and the results are shown below in Table 3, 4, 5, 6 and Fig 2, 3, 4, 5 for binary classification and Table 7, 8, 9, 10 and Fig 6, 7, 8, 9 for multiclass classification.

Table 3. Accuracy for Binary Classification

Attacks	Methods	GNB	BNB	DT	KNN	LR	SVM	RF	SGD
DDoS	F 1	0.790294	0.925517	0.999844	0.999645	0.989124	0.998518	0.999824	0.987702
	F 2	0.881810	0.970194	0.999873	0.999725	0.997447	0.998776	**0.999888**	0.997806
PortScan	F 1	0.994314	0.994896	0.999817	0.999750	0.995794	0.998454	0.999857	0.996359
	F 2	0.994697	0.993733	0.999850	0.999800	0.996924	0.999235	**0.999867**	0.999035
Infiltration	F 1	0.898642	0.979754	0.999901	0.999350	0.999800	0.999818	0.999901	0.999818
	F 2	0.908904	0.995462	0.999810	0.999825	0.999818	0.999818	**0.999837**	0.999817
Botnet	F 1	0.897649	0.414338	0.999601	0.998703	0.994291	0.992945	0.999077	0.989630
	F 2	0.995787	0.791659	0.999810	0.999077	0.994400	0.993793	**0.999845**	0.991375

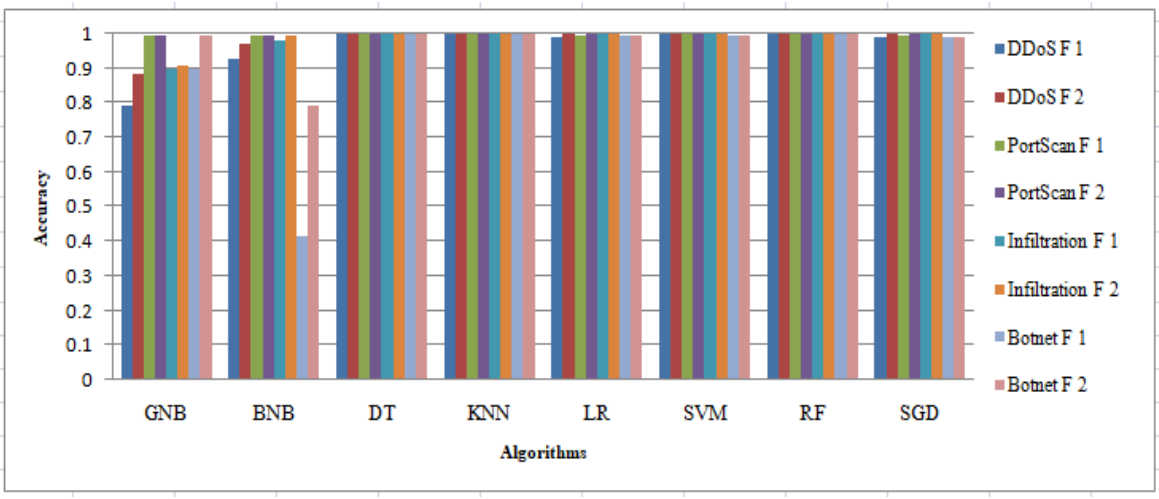

Fig 2. Accuracy for Binary Classification

Table 4. Precision for Binary Classification

Attacks	Methods	GNB	BNB	DT	KNN	LR	SVM	RF	SGD
DDoS	**F 1**	0.966263	0.884782	0.999809	0.999556	0.996163	0.998742	0.999850	0.980063
	F 2	0.972054	0.997430	0.999814	0.999667	0.996381	0.999116	**0.999862**	0.997430
PortScan	**F 1**	0.993835	0.988755	0.999788	0.999758	0.993669	0.998462	0.999751	0.988755
	F 2	0.997471	0.994420	0.999819	0.999849	0.995224	0.999037	**0.999879**	0.994239
Infiltration	**F 1**	0.001788	0.038732	0.888888	0.996775	0.998978	0.757357	0.657834	0.993765
	F 2	0.001989	0.967462	0.818181	0.998896	0.999675	0.999734	**0.999800**	0.998867
Botnet	**F 1**	0.079918	0.017312	0.980769	0.906250	0.785932	0.668354	0.951923	0.005474
	F 2	0.976833	0.979840	0.983091	0.935632	0.796296	0.996002	**0.999806**	0.999276

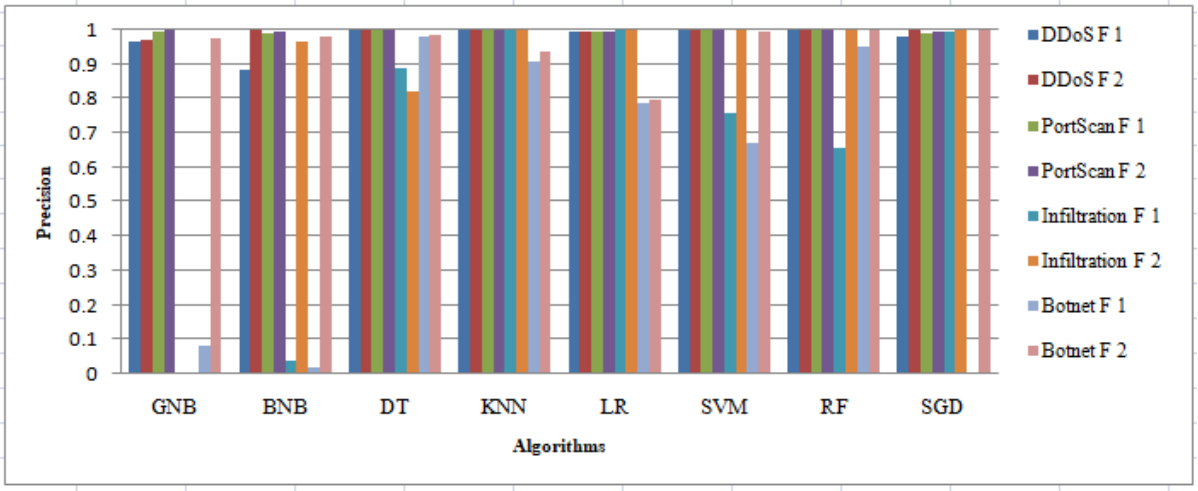

Fig 3. Precision for Binary Classification

Table 5. Recall for Binary Classification

Attacks	Methods	GNB	BNB	DT	KNN	LR	SVM	RF	SGD
DDoS	**F 1**	0.638072	0.933088	0.999762	0.999962	0.998593	0.998135	0.999901	0.933088
	F 2	0.816112	0.999481	0.999762	0.999762	0.999370	0.999111	**0.999825**	0.998741
PortScan	**F 1**	0.994210	0.998748	0.998404	0.999788	0.998733	0.998733	0.998745	0.980736
	F 2	0.996562	0.994210	0.999835	0.999788	0.999215	0.999259	**0.999851**	0.999185
Infiltration	**F 1**	0.846153	0.846153	0.615384	0.230769	0.153846	0.153846	0.615384	0.153846
	F 2	0.846153	0.996742	0.692307	0.230769	0.153846	0.995632	**0.998767**	0.979783
Botnet	**F 1**	0.608173	0.789526	0.970769	0.975961	0.617788	0.634615	0.981923	0.789526
	F 2	0.843705	0.959192	0.978365	0.978365	0.620192	0.997733	**0.998481**	0.815835

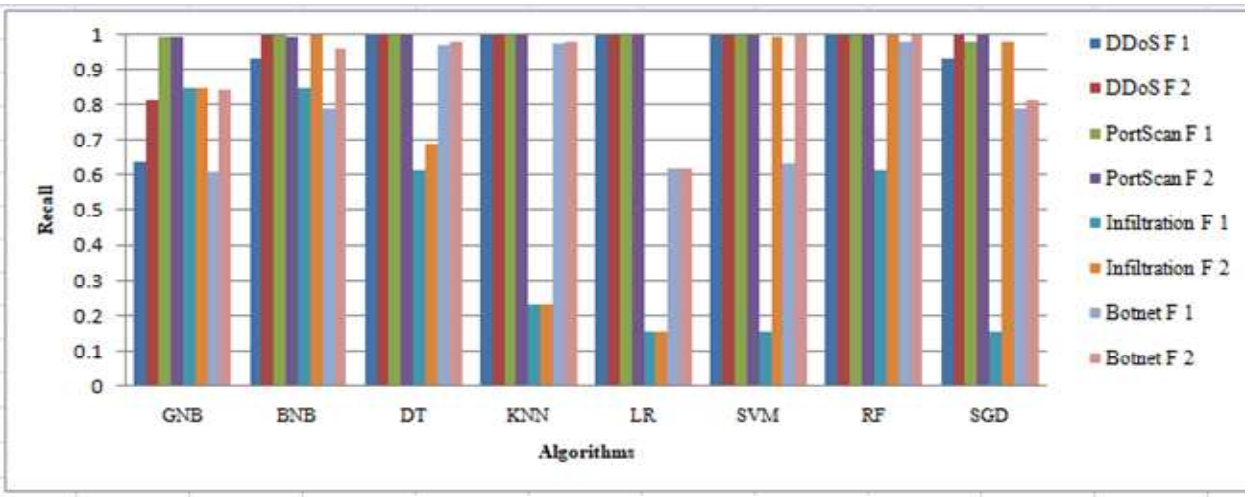

Fig 4. Recall for Binary Classification

Table 6. F1Score for Binary Classification

Attacks	Methods	GNB	BNB	DT	KNN	LR	SVM	RF	SGD
DDoS	**F 1**	0.768599	0.908293	0.999785	0.999758	0.997376	0.998485	0.999819	0.955998
	F 2	0.887283	0.908362	0.999843	0.999814	0.997773	0.999114	**0.999897**	0.998085
PortScan	**F 1**	0.994022	0.993726	0.999095	0.999707	0.996194	0.998597	0.999247	0.984729
	F 2	0.997016	0.994314	0.999824	0.999818	0.999215	0.999147	**0.999864**	0.996707
Infiltration	**F 1**	0.003568	0.074073	0.727272	0.374772	0.266699	0.255741	0.635901	0.266443
	F 2	0.003968	0.981883	0.749999	0.374946	0.266854	0.997678	**0.999283**	0.989232
Botnet	**F 1**	0.141271	0.033881	0.975743	0.939814	0.691807	0.651047	0.966690	0.010872
	F 2	0.905401	0.969406	0.980722	0.956531	0.797842	0.997066	**0.999143**	0.898285

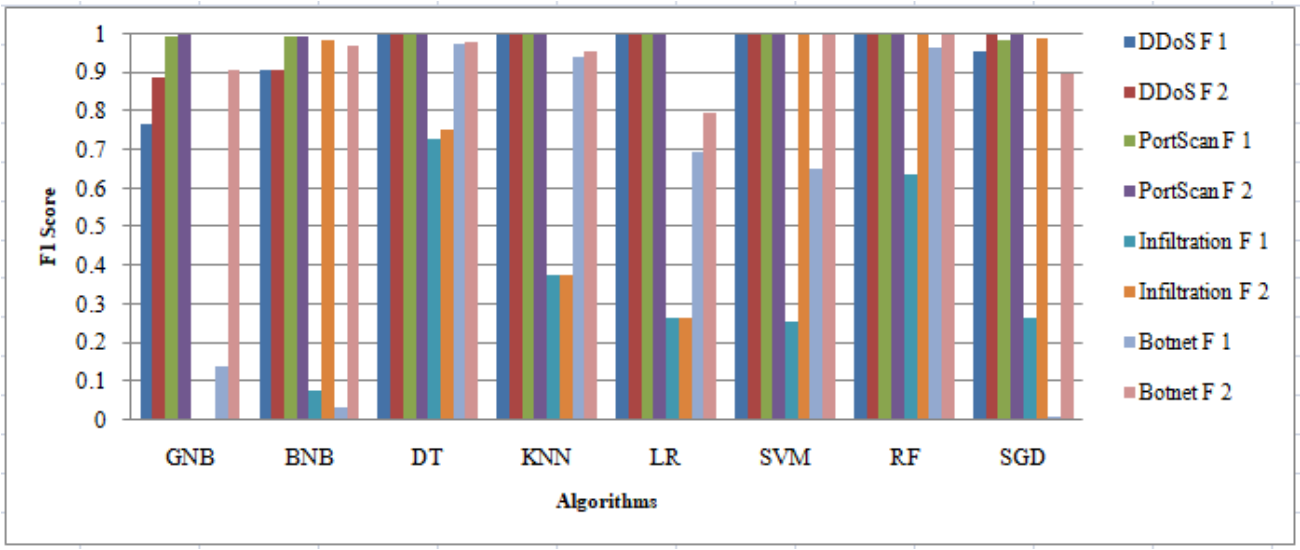

Fig 5. F1 Score for Binary Classification

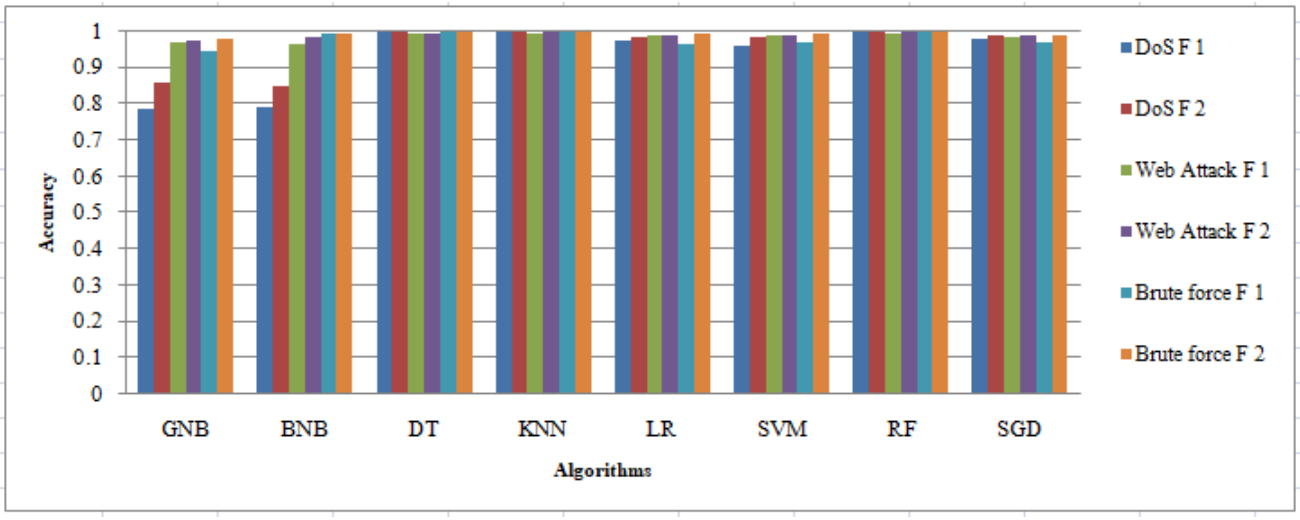

Fig 6. Accuracy for Multiclass Classification

Table 7. Average Accuracy for Multiclass Classification

Attacks	Methods	GNB	BNB	DT	KNN	LR	SVM	RF	SGD
DoS	**F 1**	0.788097	0.788654	0.999490	0.998989	0.974296	0.957835	0.999319	0.979665
	F 2	0.859371	0.847384	0.999660	0.998996	0.984594	0.984737	**0.999787**	0.989636
Web Attack	**F 1**	0.967045	0.966207	0.995440	0.995667	0.986332	0.987701	0.995527	0.984906
	F 2	0.973865	0.984235	**0.995527**	0.999524	0.990776	0.989769	**0.999840**	0.987394
Brute force	**F 1**	0.947138	0.992802	0.999834	0.999850	0.965314	0.969660	0.999857	0.969682
	F 2	0.978153	0.992855	0.999867	0.999881	0.992471	0.992492	**0.999901**	0.986352

Table 8. Average Precision for Multiclass Classification

Attacks	Methods	GNB	BNB	DT	KNN	LR	SVM	RF	SGD
DoS	F 1	0.567256	0.809837	0.638642	0.785373	0.967368	0.783462	0.738573	0.647847
	F 2	0.906173	0.997722	0.999245	0.997984	0.942425	0.986583	**0.999853**	0.987634
Web Attack	F 1	0.053376	0.268337	0.716329	0.719300	0.568728	0.356874	0.557987	0.568343
	F 2	0.304523	0.787878	0.986842	0.969696	0.193548	0.977725	**0.999745**	0.968694
Brute force	F 1	0.411016	0.997722	0.998779	0.997748	0.968170	0.969111	0.873672	0.827856
	F 2	0.697058	0.998589	0.995783	0.994574	0.262370	0.969712	**0.998756**	0.969713

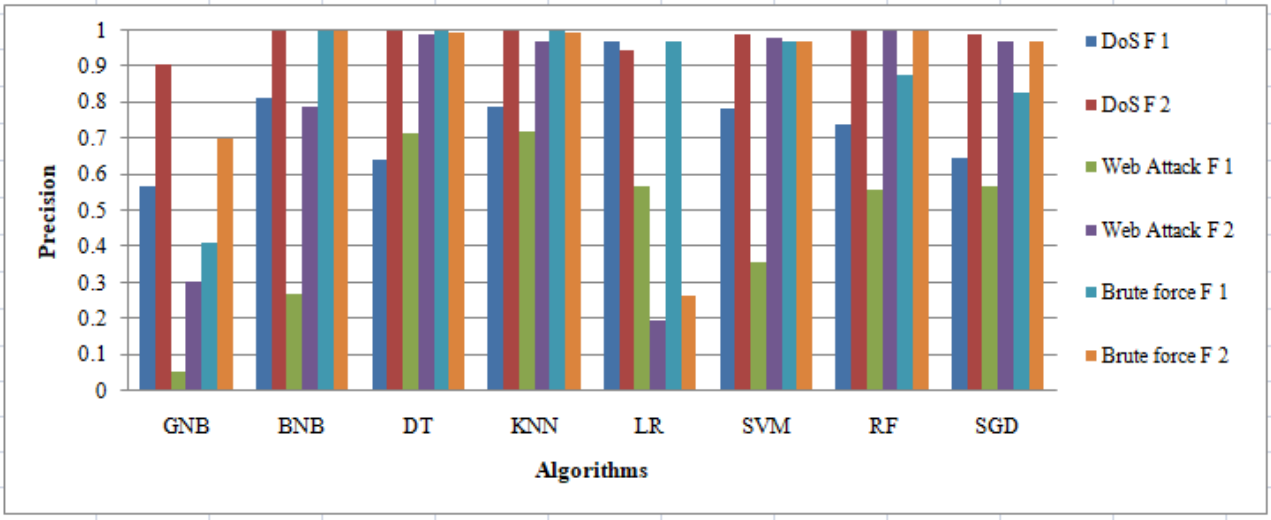

Fig 7. Precision for Multiclass Classification

Table 9. Average Recall for Multiclass Classification

Attacks	Methods	GNB	BNB	DT	KNN	LR	SVM	RF	SGD
DoS	F 1	0.685027	0.763625	0.864724	0.567295	0.406726	0.848423	0.767353	0.775792
	F 2	0.764854	0.926485	0.998763	0.999245	0.989945	0.965743	**0.999811**	0.986537
Web Attack	F 1	0.315430	0.080246	0.721923	0.490474	0.026607	0.679842	0.556473	0.428254
	F 2	0.835920	0.973392	0.997782	0.992348	0.336021	0.999771	**0.999843**	0.957373
Brute force	F 1	0.750452	0.995691	0.992373	0.996131	0.030637	0.995145	0.999254	0.747572
	F 2	0.506410	0.998589	0.999084	0.999389	0.747738	0.998544	**0.999896**	0.987473

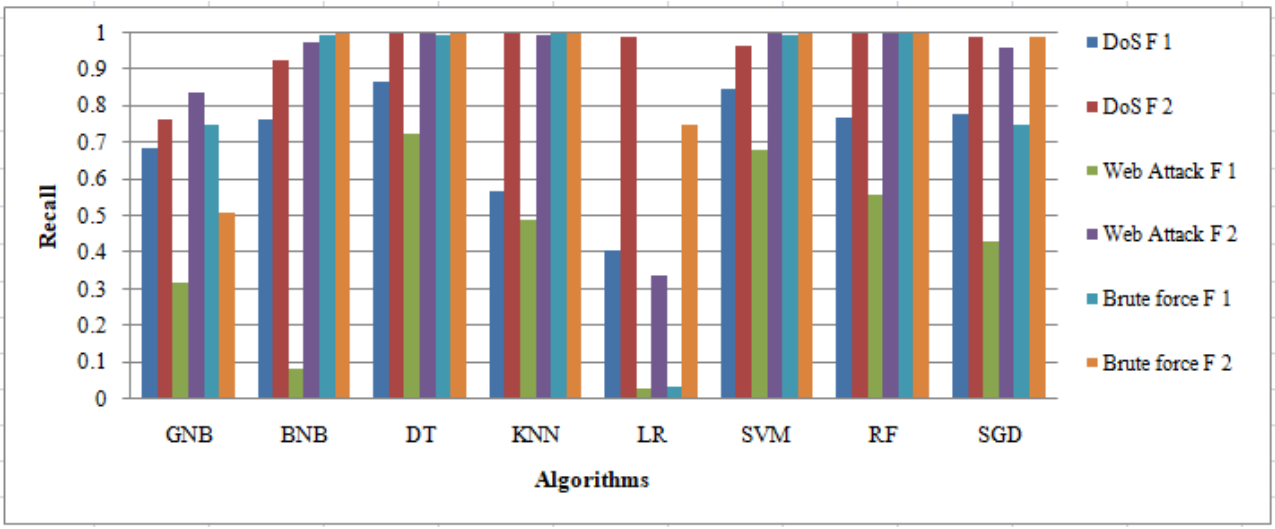

Fig 8. Recall for Multiclass Classification

Table 10. Average F1 Score for Multiclass Classification

Attacks	Methods	GNB	BNB	DT	KNN	LR	SVM	RF	SGD
DoS	F 1	0.620603	0.785552	0.734683	0.658754	0.573090	0.818649	0.752683	0.706070
	F 2	0.829537	0.960784	0.999307	0.998610	0.965600	0.976051	**0.999527**	0.987088
Web Attack	F 1	0.091302	0.123545	0.719115	0.583246	0.050835	0.468050	0.557228	0.488452
	F 2	0.449220	0.870864	0.992281	0.980891	0.245619	0.988625	**0.999793**	0.963000
Brute force	F 1	0.531134	0.996705	0.995565	0.996934	0.059394	0.981925	0.932252	0.785668
	F 2	0.586633	0.998589	0.997430	0.996975	0.383441	0.983916	**0.999325**	0.978512

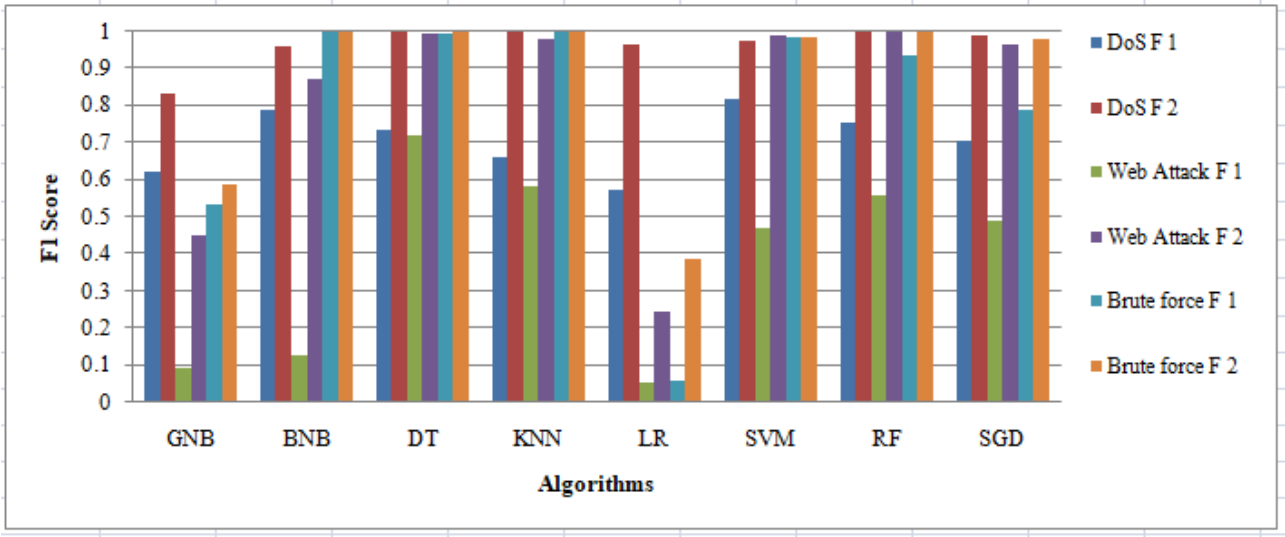

Fig 9. F1 Score for Multiclass Classification

We have calculated the above evaluation parameter in two categories:

1. Select all Dataset without Data Preprocessing and Feature Selection (F1) for Multiclass classification and Binary classification.
2. With Data Preprocessing and Feature Selection (F2) for Multiclass classification and Binary classification.

From both evaluation methods, we found that when we used feature selection with data pre-processing for each attack (DoS, PortScan, Botnet, Web Attacks, Infiltration, Heartbleed and, DDoS) on the dataset, for various algorithms (GaussianNB (GNB), BernoulliNB (BNB), Decision Tree, KNN, Logistic Regression, SVM, SGD, and Random Forest) evaluation parameters are increasing. This means that data pre-processing and feature selection are an important task for detecting attacks on the dataset.

Feature selection removes unwanted and redundant features and improves the efficiency of the machine learning model. When applying various ML techniques in Binary Classification, we observed that (RF) Random Forest algorithm with Data pre-processing and Feature Selection is gives highest accuracy (0.999888) when compared with other machine learning techniques (Table 3 and Fig 2). Apart from this, when comparing precision (0.999800), F1 Score (0.999897) and recall (0.999851) parameters with other machine learning algorithms, we observed that Random Forest algorithm gives the best results (Table 4, 5, 6, and Fig 3, 4, 5). Also when calculating and analysing the results in multiclass classification for different attacks on the dataset, we have seen that the Random Forest algorithm with Data pre-processing and Feature Selection gives highest accuracy (0.999901) (Table 7 and Fig 6).

Table 11. Comparison of Our Results with Past Studies Based on Four Evaluation Criteria

Machine Learning Algorithms	Sharafaldin et al. 2018 [8].				Kahraman Kostas 2018 [25]			
	Recall	F1 Score	Precision	Accuracy	Recall	F1 Score	Precision	Accuracy
Random Forest	0.97	0.97	0.98	---------	0.94	0.94	0.94	0.94
Naïve Bayes	0.84	0.84	0.88	---------	0.87	0.86	0.86	0.86
KNN	0.96	0.97	0.96	---------	0.97	0.97	0.97	0.97

Table 12. Comparison of Our Results with Past Studies Based on Four Evaluation Criteria for Binary Classification

Machine Learning Algorithms	Vinayakumar et al. 2019 [4]			
	Recall	F1 Score	Precision	Accuracy
Random Forest	0.969	0.905	0.849	0.940
Naïve Bayes	0.979	0.459	0.300	0.313
KNN	0.968	0.865	0.781	0.910
Logistic Regression	0.850	0.758	0.685	0.839
Decision Tree	0.965	0.898	0.839	0.935
SVM	0.328	0.493	0.992	0.799

Table 13. Comparison of Our Results with Past Studies Based on Four Evaluation Criteria for Multiclass Classification

Machine Learning Algorithms	Vinayakumar et al. 2019 [4]			
	Recall	F1 Score	Precision	Accuracy
Random Forest	0.944	0.953	0.970	0.944
Naïve Bayes	0.250	0.188	0.767	0.250
KNN	0.909	0.922	0.949	0.909
Logistic Regression	0.870	0.868	0.889	0.870
Decision Tree	0.940	0.949	0.965	0.940
SVM	0.915	0.723	0.757	0.915

And the best results in term of precision (0.999853), recall (0.999896) and F1 Score (0.999793) (Table 8, 9, 10, and Fig 7, 8, 9).

Table 11, 12 and 13 shows the results obtained from the previous studies. When we compare these results with our results in Table no 3 to 10, then we see that our results are better than in terms of F1 score, recall, precision and accuracy.

V. Conclusion

The purpose of this study is to build an intrusion detection system using different machine learning techniques to detect anomaly or intrusion on the CICIDS-2017 dataset. The CICIDS- 2017 dataset has been used due to a variety of well-known updated attacks (Binary or Multiclass Classification) in it. We have proposed a model in which two approaches have been used to create efficient intrusion detection system. In first step, we did data pre-processing including data cleaning, one-hot encoding, etc. In the second step, we did a feature selection that is features that are not necessary, redundant, or not correlated to other attributes are eliminated from the dataset.Finally, eight ML techniques are used for intrusion detection, which are significantly used and have distinct nature and characteristics. Specifically, we have analyzed eight machine learning techniques (GaussianNB (GNB), BernoulliNB (BNB), Random Forest, KNN, Logistic Regression, SVM, SGD, and Decision Tree). These ML approaches are compared using the CICIDS-2017 dataset with four important performance indicators namely, accuracy, recall, precision, and F1 Score.

In the future, this research will work towards developing an efficient intrusion detection system (IDSs) to detect intrusion based on the investigation.

Furthermore, we have already started using this approach on the CIDIDS 2017 dataset with the help of deep learning algorithms and hybrid machine learning algorithms to achieve a higher level of performance.

References

1. S. Rodda, Network Intrusion Detection Systems Using Neural Networks, Information Systems Design and Intelligent Applications Advances in Intelligent Systems and Computing in Springer Nature Singapore, vol. 672, 2018.
2. A. H. Ngu, M. Gutierrez, V. Metsis, S. Nepal, Q. Z. Sheng, IoT middleware: A survey on issues and enabling technologies, IEEE Internet of Things Journal, vol. 4, no. 1, pp. 1–20, Feb 2017.
3. W. Fu, X. Xin, P. Guo, and Z. Zhou, A practical intrusion detection system for internet of vehicles, China Communications, vol. 13, no. 10, pp. 263–275, Oct 2016.
4. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, S. Venkatraman, Deep Learning Approach for Intelligent Intrusion Detection System, in IEEE Access, vol. 7, pp. 41525-41550, 2019.
5. S.M. Othman, Ba-Alwi, F.M., Alsohybe, N.T. et al. Intrusion detection model using machine learning algorithm on Big Data environment. Journal of Big Data, vol. 5, article no. 34, 2018.
6. K. Wu, Z. Chen, W. Li, A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks, in IEEE Access, vol. 6, pp. 50850-50859, 2018.
7. R. Abdulhammed, H. Musafer, A. Alessa, M. Faezipour, A. Abuzneid, Features Dimensionality Reduction Approaches for Machine Learning Based Network Intrusion Detection, in MDPI Electronics, vol. 8, Issue 3, 2019.
8. I. Sharafaldin, A. H. Lashkari, A. A. Ghorbani, Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization, 4th International Conference on Information Systems Security and Privacy (ICISSP), Purtogal, January 2018.
9. T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, M. Ghogho, Deep learning approach for Network Intrusion Detection in Software Defined Networking, *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Fez, pp. 258-263, 2016.
10. J. Kim, J. Kim, H. L. Thi Thu, H. Kim, Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection, 2016 International Conference on Platform Technology and Service (PlatCon), Jeju, pp. 1-5, 2016.

11. Q. Niyaz, W. Sun, A. Y Javaid, M. Alam, Deep Learning Approach for Network Intrusion Detection System, Endorsed Transactions on Ambient Systems, vol 3, 2016.
12. C. Yin, Y. Zhu, J. Fei, X. He, A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks, in IEEE Access, vol. 5, pp. 21954-21961, 2017.
13. N. Marir, H. Wang, G. Feng, B. Li, M. Jia, Distributed Abnormal Behavior Detection Approach Based on Deep Belief Network and Ensemble SVM Using Spark, in IEEE Access, vol. 6, pp. 59657-59671, 2018.
14. Z. Zhang, X. Zhou, X. Zhang, L. Wang, Pengwei Wang, A Model Based on Convolutional Neural Network for Online Transaction Fraud Detection, Security and Communication Networks, vol 2018.
15. T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, M. Ghogho, Deep Recurrent Neural Network for Intrusion Detection in SDN-based Networks, 4th IEEE Conference on Network Softwarization and Workshops (NetSoft), Montreal, QC, 2018, pp. 202-206, 2018.
16. F. Jiang et al., Y. Fu , B. B. Gupta Y. Liang, S.Rho, F. Lou, F.Meng, Z. Tian, "Deep Learning Based Multi-Channel Intelligent Attack Detection for Data Security," in IEEE Transactions on Sustainable Computing, vol. 5, no. 2, pp. 204-212, 1 April-June 2020..
17. D. Aksu, S. Üstebay, M.A. Aydin, T. Atmaca T, (2018) Intrusion Detection with Comparative Analysis of Supervised Learning Techniques and Fisher Score Feature Selection Algorithm, in Computer and Information Sciences, ISCIS, Springer, pp. 141–149, vol 935, 2018.
18. N. Acharya, S. Singh, An IWD-based feature selection method for intrusion detection system, Soft Computing, vol 22, pp. 4407–4416, 2018.
19. U. Cavusoglu, A new hybrid approach for intrusion detection using machine learning methods. Applied Intelligence, vol. 49, pp. 2735–2761, 2019.
20. P. Negandhi, Y. Trivedi, R. Mangrulkar, Intrusion detection system using random forest on the NSL–KDD dataset, Emerging research in computing Information communication and applications. Springer, Berlin, pp 519–531, 2019.
21. S. S. Panwar, Y. P. Raiwani, L. P. Panwar, Evaluation of Network Intrusion Detection with Features Selection and Machine Learning Algorithms on CICIDS-2017 Dataset, International Conference on Advances in Engineering Science Management & Technology (ICAESMT) - 2019, Uttaranchal University, Dehradun, India, March 2019 Available at SSRN: https://ssrn.com/abstract=3394103
22. S. S. Panwar, P. S. Negi, , L. P. Panwar, Y. P. Raiwani , Implementation of Machine Learning Algorithms on CICIDS-2017 Dataset for Intrusion Detection Using WEKA, International Journal of Recent Technology and Engineering (IJRTE), vol. -8, Issue-3, September 2019.
23. V. Dutta, M. Choras, R. Kozik, M. Pawlicki, Hybrid Model for Improving the Classification Effectiveness of Network Intrusion Detection. 13th International Conference on Computational Intelligence in Security for Information Systems (CISIS), Advances in Intelligent Systems and Computing, vol 1267. Springer, Cham. 2020. .
24. M. Mok, S. Sohn, Y. Ju, Random effects logistic regression model for anomaly detection, *Expert System Applied,* vol. 37, no. 10, pp. 7162-7166, 2010.
25. K. Kostas, Anomaly Detection in Networks Using Machine Learning, School of Computer Science and Electronic Engineering University of Essex, August 2018.