ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume **23**, Number 3, 2020, 250–261

Evaluating the Impact of Feature Selection Methods on the Performance of the Machine Learning Models in Detecting DDoS Attacks

Naveen BINDRA and Manu SOOD

Himachal Pradesh University, Shimla, India
E-mail: naveenjb@hotmail.com, soodm_67@yahoo.com

Abstract. Heaps of Data lie in network equipment of the organizations. To break down this information and reach some significant inferences is inconceivable for the present day IDS (*Intrusion Detection System*). Moreover, their signature-based defense mechanisms are ineffective in tackling emerging threats like DDoS attacks. The central premise of building Machine Learning Classifiers is to detect DDoS attacks efficiently and effectively. However, their accomplishment of machine learning models to distinguish DDoS attacks relies upon how one picks the 'relevant' and 'minimal' attributes in the network streams. The questions: "Does features choice influence the classification precision, to what extent and what is the best feature selection for boosting the performance of Machine Learning Classifiers in detecting DDoS attacks" motivate this investigation.

This paper presents how to apply ML techniques in detecting DDoS attacks. There are three feature selection categories and our work demonstrates the application of the feature selection methods, one each from the three categories. In addition, five machine learning models have been utilized in our work. Random Forest classifier with Lasso-RFE, a feature selection strategy beat others. The other significance of this research paper, we believe, is a first of its kind scenario, where a real-life multidimensional dataset having diverse network traffic and recent DDoS attacks have been used unlike in most of the studies carried out to date.

Key-words: DDoS detection, DDoS attack, machine learning, feature selection, SciKit-Learn, network traffic classification.

1. Introduction

Keeping computer network safe is a task for organizations. Complex and mutating DDoS attacks are the primary vectors for causing damage to the networks, data and, availability of services. With the development of Big Data, IoT (Internet of things) and BYOD (*Bring Your Own*

Device), and social networking applications, the network traffic has increased many folds and so are the network threats [1][2][3]. Traditional Intrusion Detection Systems are not that shrewd to recognize advanced attacks. Therefore, there is a need to devise new strategies/arrangement, which can go up against these threats. The other real obstacle in utilizing existing strategies is 'human intervention', which is required at present for the elucidation of threat recognition. With the establishment of a large campus-wide area network, internet facilities, Wi-Fi, attack-surfaces have multiplied.

Heaps of Data lie with any organizations' network equipment such as switches, firewalls, controllers and so forth. To break down this information and reach some significant inferences is inconceivable for the present day IDS (Intrusion Detection System) frameworks because of vast information estimate. Moreover, their signature-based defense mechanisms are ineffective in tackling emerging threats. Thus, analyzing the data and draw some meaningful conclusions is next to impossible for the present day IDS (Intrusion Detection System) systems. Filtering the relevant data requires a huge amount of human effort.

The role of machine learning in intrusion detection particularly DDoS attacks are now well established. Machine learning (ML) and the information mining techniques, together are exceptionally viable in such circumstances. The use of ML-based classifiers in getting rid of the need for human mediation is a potent way of recognizing inconspicuous attacks, which is yet odd for other security machines. The ML-based classifiers learn on the training data and anticipate the Intrusion identification and DoS attacks. The success of classifiers vastly depends upon relevant features and classification accuracy. Classification accuracy is the ratio of the number of correct predictions to the total number of input samples [4]. Feature selection helps in retaining only the meaningful features and drops the redundant and correlated features in order to boost prediction accuracy reduce computation complexities and prediction latency. To train the machine learning models attaining adequate accuracy need extensive data. Huge data generally has a noise that affects the performance of machine learning models adversely. Thus, the data need cleansing and removal of redundant and irrelevant features for machine learning models to make predictions with high accuracy. Redundant features are those, which correlate strongly, and removal of them does not affect the accuracy of the model. Similarly, many features in the dataset, which does not influence the prediction of machine learning models are termed as 'irrelevant' and are dropped during feature selection. Thus, 'noise removal' plays an important role in building ML models.

Optimizing the performance of feature selection techniques is an optimizing problem in mathematical terms. Feature selection has a three-fold impact on the overall performance of machine learning models: 1) increasing the accuracy of machine learning models, 2) helps in reducing the performance latency of various ML algorithms and 3) diminish the computational complexity to analyze their correlation.

This examination embarks to explore the aptness of machine learning in network traffic classification by using the results obtained from various feature selection methods. The contribution of this work can be summarized as below:

- We present a study of detecting DDoS attacks in an extensive dataset with diverse normal traffic and the latest DDoS attacks unlike other studies carried out to date.

- We present the study over the CICIDS2017 dataset, which is a real-life dataset with over 225,000 rows and over 84 features. To the best of our knowledge, this study is the first one where this contemporary dataset has been put to use. We also show the impact of feature selection and we were successful in achieving an accuracy above 95 and 96% using

just 12/15 features in case of KNN and Random Forest ML models respectively, without compromising on the accuracy.

This paper organizes into five sections. Section II outlines the work done by researchers in the literature to explore network flow classification. Section III of this paper gives an overview of the dataset chosen for this experiment, feature selection methods and classifiers used for the detection of DDoS. Section IV describes our experiment and the selection of tools for analysis. Of all the sections, section IV is the most vital section as it divulges details about the results and the manner they are obtained. Section V wraps up the work and also indicates the avenue for future work.

2. Related Work

Feature selection has an imperative impact in boosting the accuracy of a model. Recent studies have revealed that selected features were successful in increasing or maintaining the accuracy of the machine learning models.

N.T. Nguyen *et al.* [5] looked at the performance of feature selection approaches namely correlation-feature-selection (CFS) measure and the *minimal-redundancy-maximal-relevance* (mRMR) measure with SVM-wrapper, Marko blanket, and *Classification & Regression Trees* (CART) algorithms as well as the recently proposed *generic-feature-selection* (GeFS) method. They used the KDD dataset and were able to remove more than 30% features while achieving better classification accuracy. However, the results are based on the KDD dataset which is outdated now for not having the current DDoS attacks. H. Nkiama *et al.* [6] demonstrate the impact of recursive feature elimination with decision tree classifier for feature selection. The classifier showed good performance over the NSL-KDD dataset. They achieved more than 99% accuracy.

Feature selection methods like IG, Gain ration and relief F, etc. along with classifiers such as J48, random and one R etc. in [7] establish the relation between classifiers and feature selection methods.

Work in [8] contrasts different feature selection strategies to discover that entirety of term frequency. This work additionally examined the combined effect of multiple feature selection models and Osanaiye et al. in [9] even propose a multi-filter feature selection method and evaluated the feature selection methods like Info gain, gain ratio, chi-squared and relief. The metrics used were classification accuracy, detection rate, false alarm rate and time to build models. Similarly, Pekta *et al.* [10] talk about identifying prominent features for the detection of botnets in the network. They examined the execution of three feature selection algorithms namely Lasso, RFE, and decision tree-based feature selection. The metrics employed for their evaluation were precision, F1 score, recall, and accuracy. The Random forest was the best performer in their case. They could accomplish 93% accuracy in the case of random forest classifier, alongside the feature selection methods.

Peyman Kabiri and Gholam Reza Zarga [11] proposed a feature-based detection of DoS attacks where their work identified the relevant features for the detection of different classes present in the dataset using PCA (*Principal Component Analysis*). The accuracy achieved by the authors for this experiment was more than 97% which is very good.

Prafulla Kalapatapu1 et al. [12] compared various feature selection methods e.g. genetic algorithm, Forward feature selection, information gain and correlation based on four classifiers

Table 1. Features in cic ids 2017 dataset

SNo	S NoFeature Name	SNo	Feature Name	SNo	Feature Name	SNo	Feature Name
1	Flow ID	22	Flow Packets/s	43	Fwd Packets/s	64	Fwd Avg Packets/Bulk
2	Source IP	23	Flow IAT Mean	44	Bwd Packets/s	65	Fwd Avg Bulk Rate
3	Source Port	24	Flow IAT Std	45	Min Packet Length	66	Bwd Avg Bytes/Bulk
4	Destination IP	25	Flow IAT Max	46	Max Packet Length	67	Bwd Avg Packets/Bulk
5	Destination Port	26	Flow IAT Min	47	Packet Length Mean	68	Bwd Avg Bulk Rate
6	Protocol	27	Fwd IAT Total	48	Packet Length Std	69	Subflow Fwd Packets
7	Timestamp	28	Fwd IAT Mean	49	Packet Length Variance	70	Subflow Fwd Bytes
8	Flow Duration	29	Fwd IAT Std	50	FIN Flag Count	71	Subflow Bwd Packets
9	Total Fwd Packets	30	Fwd IAT Max	51	SYN Flag Count	72	Subflow Bwd Bytes
10	Total Backward Packets	31	Fwd IAT Min	52	RST Flag Count	73	Init_Win_bytes_forward
11	Total Length of Fwd Packets	32	Bwd IAT Total	53	PSH Flag Count	74	Init_Win_bytes_backward
12	Total Length of Bwd Packets	33	Bwd IAT Mean	54	ACK Flag Count	75	act_data_pkt_fwd
13	Fwd Packet Length Max	34	Bwd IAT Std	55	URG Flag Count	76	min_seg_size_forward
14	Fwd Packet Length Min	35	Bwd IAT Max	56	CWE Flag Count	77	Active Mean
15	Fwd Packet Length Mean	36	Bwd IAT Min	57	ECE Flag Count	78	Active Std
16	Fwd Packet Length Std	37	Fwd PSH Flags	58	Down/Up Ratio	79	Active Max
17	Bwd Packet Length Max	38	Bwd PSH Flags	59	Average Packet Size	80	Active Min
18	Bwd Packet Length Min	39	Fwd URG Flags	60	Avg Fwd Segment Size	81	Idle Mean
19	Bwd Packet Length Mean	40	Bwd URG Flags	61	Avg Bwd Segment Size	82	Idle Std
20	Bwd Packet Length Std	41	Fwd Header Length	62	Fwd Header Length	83	Idle Max
21	Flow Bytes/s	42	Bwd Header Length	63	Fwd Avg Bytes/Bulk	84	Idle Min

(Decision tree C4.5, K-Nearest neighbors, neural network and support vector machine. They claim to achieve higher accuracy with these feature selection methods than otherwise.

This work is better on three grounds from similar works. First, the use of the CICIDS2017 data set, which has an edge over other datasets in terms of being realistic and entailing contemporary DDoS attacks. Secondly, it covers all the three categories of feature selection and finally, the feature selection, in our case we were able to remove 82% feature which is well below the works considered here.

3. Data Pre-processing and Feature selection

A. Dataset

Among the available choices for the IDS dataset, we picked CIC IDS 2017 dataset [13]. The dataset includes contemporary DoS attacks and benign network flows to imitate the real scenario [14]. The attacks used for the generation of data are Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet, and DDoS. The dataset involves about eleven criteria such as attack diversity, labeling, complete capture, complete interaction etc. the experimental setup was comprised of physical machines and tools like curl have been used to generate normal network traffic. The network topology used for capturing the network traffic includes most of the network devices and the dataset is complete capture. Network protocols *e.g.* HTTP, HTTPS, SSH, FTP, SMTP, IMAP, and POP3 have been used for the generated traffic to make it more real. They capture data for five days and we chose data captured on Friday for our experiments, which included dos attacks and benign traffic. The dataset runs into GigaByte and was provided as labeled data as a '.csv' file format. With about 85 features as mentioned in Table I, the dataset is an ideal choice for demonstrating the effect of feature selection on how five Machine Learning based classifiers are chosen for this study perform.

B. Feature selection

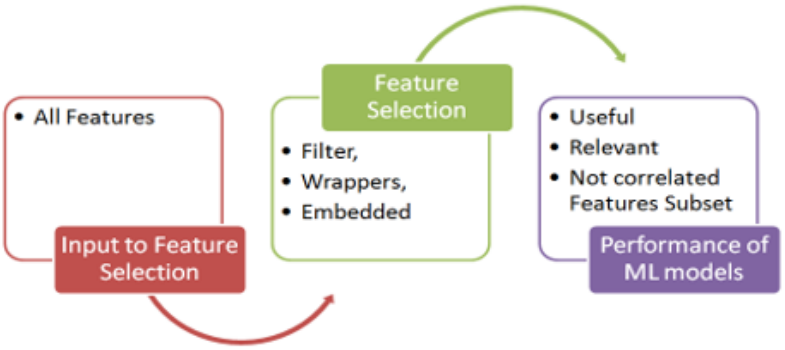

Fig. 1. Feature Selection Process.

Feature selection [15][22] as the name suggests, is the selection of the features out of the total available attributes or features in order to reduce the computational latency and complexity, increasing accuracy and doing away with 'overfitting'. It is three types 1) Filter methods, 2) Wrappers Methods and 3) embedded methods. In filter methods, correlation of the features with the dependent variable ascertains their relevance while wrapper methods use the actual classifier to weigh the worth in subset of the attributes. Filter methods are much faster and less computationally intensive as compared to the wrapper methods as they do not involve training the models. Feature selection methods [15] categorize into Filter, Wrappers, and Embedded Methods.

Recursive Feature Elimination [16] is an embedded method that works with machine learning algorithms like Support Vector Machines, Lasso, etc. Here the model construction happens by repeated removal of the features with low weights. Embedded methods have computational complexity between wrappers and filter methods.

Selectfrommodel [17] strategy is more or less the same as RFE but is less robust. Here the features are chosen in view of weights and as per the threshold. The estimator is a parameter for this method which is machine learning algorithms like Random Forest, lassoCV, etc.

SelectPercentile [18] is another feature selection method that can be implemented using Scikit-learn which assigns percentile to the features based on their score. Subsequently, features can be selected to a cutoff percentile considering the performance of a classifier.

PCA (Principal Component Analysis) [19] projects correlated features on a different plane making them uncorrelated i.e. with maximum variance. In mathematical terms, it can also be defined as an orthogonal linear transformation. For a given set of k-dimensional vectors of weights w(k) , there exists map x(i) from X to orthogonal vector Y(i) is represented by Y(i) = x(i).w(k). In PCA, the dimensionality of data is reduced but the original features are not retained by this method unlike in other feature selection methods.

C. Classifiers

A classifier in machine learning is a statistical function that classifies network traffic. We used five supervised ML-based classifiers for our study. These are Support Vector Machine (SVM), Gaussian Nave Bays (GNB), and K nearest neighbor and Random Forest. Support Vector Machine [20,21] is a supervised machine learning algorithm that finds a line of categories by

learning from the dataset. The major benefit of using this classifier is that it doesn't 'overfit'. The classification is carried out by searching for a hyperplane to classify the classes in the best possible way. We have used Linear SVM which is a type of SVM and is very efficient. Nave Bayes Classifier [20,21] is amongst the most prominent learning methods based on Bayes Theorem of Probability. It is a very fast method and is capable of making almost real-time predictions. The main downside of this method is its supposition regarding independent predictors which is not possible practically. Logistic regression [20,21] is a statistical method for analyzing the dataset having one or more independent variables for determining the outcome. Random Forest [20, 21] is the group learning method used for classification. Random Forest is a group learning method used for classification. It creates several decision trees during its training and uses votes from these decision trees for predicting an output. K-nearest neighbor algorithm [20, 21] is generally used for both classification and regression. This is also known as the lazy learning algorithm. It stores every single accessible case and predicts the class of the information in view of these cases.

4. Experiments & Results

A. Testbed

The experiments were carried out over Ubuntu (64-bit) based virtual machine having windows 10 as host. The memory of the physical machine was 4 GB. With a Core 2 Duo processor, it was a robust machine in the business segment and was tuned for best performance i.e. we disabled most of the graphics/ redundant processes running in the background. The software application was written by the authors in Python programming language using Scikit-learn Library [20]. The pseudo-code of the same is given in Fig. 2.

B. Metrics

We used five machine learning models and as many feature selection methods. The accuracy of the classifiers and number of the features selected by the feature selection methods without compromising on the accuracy of the classifier were utilized as metrics for the evaluation of our experiments. The accuracy of Machine Learning classifiers can be calculated as below:

Accuracy = correct predictions/ total number of instances

C. Outcome

We ran the experiment initially without pre-processed data; the results are illustrated in Fig. 3 & Table I. The drawback observed was the higher computational latency. The dataset was subjected to various data pre-processing techniques such as removing blank values and replacing NaN with median values and Infinity values with '0'. 'Infinity' values are very large values and NaN is 'Not a Number' generally creep in instead of the intended observation when something goes wrong during the happening of an experiment. It was logical to replace/drop these values with appropriate methods referred above.

A software module was written to compare the feature selection techniques. The same was developed using 'python' and python based libraries such as Scikit-learn [20], Numpy and Pandas, etc. The pseudo-code snippet in Fig. 2 depicts the logic of the code and Fig. 3 is the screenshot of the actual results obtained during the trial. We used 10 fold cross-validation, which

is a far advanced technique than 'train_test_split' strategy [23]. It divides the dataset into 10 segments of equal size and each segment is used iteratively for testing and training of the model. We picked classification accuracy as the metric to evaluate the performance of feature selection methods.

```
IMPORT important libraries including Scikit Learn
IMPORT the dataset
PRE-PROCESSING to impute missing values, replace NaN (Not a
Number) and Infinity values in the dataset
DROP redundant/ irrelavant features
SCALE the data
STOREvarious Machine Learning Models in a variable 'models'
SET scoring equal to accuracy, Name as name of the machine learning
models
FOR Name, Model in models:
      Store value of model  selection using 10 splits in a variable
       Calculate and store results using cross_val_score method of
        model_selection in sklearn by imputing Train and Testing data
        Append results in list of existing results
        Print mean accuracy and standard deviation
END FOR
```

Fig. 2. Pseudo Code snippet for DDoS traffic classifier.

It was observed that there was a clear difference in execution time in the two scenarios *i.e.* with and without scaling. Without scaling of data, the classifiers took longer to finish. However, after applying feature selection methods, the time taken by the model to finish was much shorter. Thus, we deduce that feature scaling and feature selection methods influence the performance of classifiers. We applied several feature selection methods *e.g.* Recursive Feature Elimination (RFE with Lasso and Linear SVC) 'SelectFromModel', 'SelectPercentile' and PCA (Principal Component Analysis). The results of the same have been illustrated in Table II.

It can easily be verified that the number of the features was reduced from 85 to 12/15 in most of feature selection methods without compromising on the accuracy of the classifier. We also replaced the categorical labels i.e. BENIGN and DDoS with '0' and '1' respectively as the machine learning models require float/ numerical values.

Apart from the application of feature selection methods, we also picked PCA for a comparison viewpoint. PCA is not a feature selection method per se but is a dimensionality reduction technique used by us to have a wider perspective of the reduced number of attributes over performance in classifying network flows in DDoS detection. The subsets selected using different feature selection methods are given in Table IV.

It can easily be verified that the number of the features was reduced from 85 to 12/15 in most of feature selection methods without compromising on the accuracy of the classifier. We also replaced the categorical labels i.e. BENIGN and DDoS with '0' and '1' respectively as the machine learning models require float/ numerical values.

Table 2. Experimental Results: Mean Accuracy of Classifiers alongside Feature Selection methods

#	Machine Learning Models	Without any feature selection with all 85 features		SS & Select Percentile (k=15)		SS & RFE (model= linearsvc and components=15)		SS & PCA(n=25)		SS and lasso CV select frommodel (threshold=0001) feature selected =12		SS with Lasso (alpha=.05) & RFE(feature to select=15)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	Logistic Regression	83.33	.268	82.4	.003	81.2	.002	82.5	.005	82.85	.300	82.53	.308
2	K Nearest Neighbour	94.12	.167	94.3	.001	80	.028	94.2	.001	95.92	.1666	95.84	.161
3	Gaussian NB	60.72	.438	81	.002	51.3	.002	51.8	.003	73.77	.263	59.82	.498
4	Random Forest	93.80	0.256	96.1	.001	82.4	.002	95.6	.001	96.5	0.170	96.65	.116
5	Linear SVM	84.16	.292	82.3	.002	82	.002	82.9	.002	83.64	0.352	83.24	.299

Fig. 3. Screen Shot of results obtained in one of the Trials.

Apart from the application of feature selection methods, we also picked PCA for a comparison viewpoint. PCA is not a feature selection method per se but is a dimensionality reduction technique used by us to have a wider perspective of the reduced number of attributes over performance in classifying network flows in DDoS detection. The subsets selected using different feature selection methods are given in Table IV.

Table 3. Feature selected with different methods

#	Feature selection methods	Feature selected
[1]	Features selected using RFE witht linearsvc and n_features_to_select =15 and step=5 (pp StandardScaler)	11, 12, 13,15, 16, 33, 41, 47, 59, 60, 70, 72, 74, 75
[2]	Feature selected after using SelectPercentile(12 features percentile=15) and using standardscaler	3, 17', 19, 20, 23, 24, 25, 29 ,30, 61, 81, 83
[3]	Lasso cv(toi=.0001) with SelectFromModel(model,threshold=-05) and StandardScaler	3, 8, 9, 23, 24', 32, 34, 47, 54', 59, 60, 61
[4]	RFE (n_features to select 15) with Lasso(alpha=.05) and StandardScaler(results reverified)	3, 5, 6, 9, 10, 11, 12, 13, 19, 24 , 54, 61, 76, 81

We were able to show the following aspects of feature selection in this study:

a) Random Forest and KNN turned out to be the best Machine Learning model to detect DDoS attacks. With feature selection methods, the Random Forest classifier was top performer and KNN performed well without using any feature selection i.e. keeping all the 85 features.

b) The classifiers perform differently with different feature classification techniques i.e. Correlation exists between classifier and feature selection algorithm.

c) The Machine Learning classifiers performed fantastically with adequate prediction accuracy over a real-life dataset which entails contemporary DDoS attacks and diverse network traffic.

d) We were successful in reducing the feature from 85 to 12/15 without compromising on the accuracy of the classifier.

e) Feature selection does boost crease the performance and accuracy of the classifier.

f) We illustrated the effect of feature selection methods belonging to all three categories e.g. Filter, wrapper, and embedded methods.

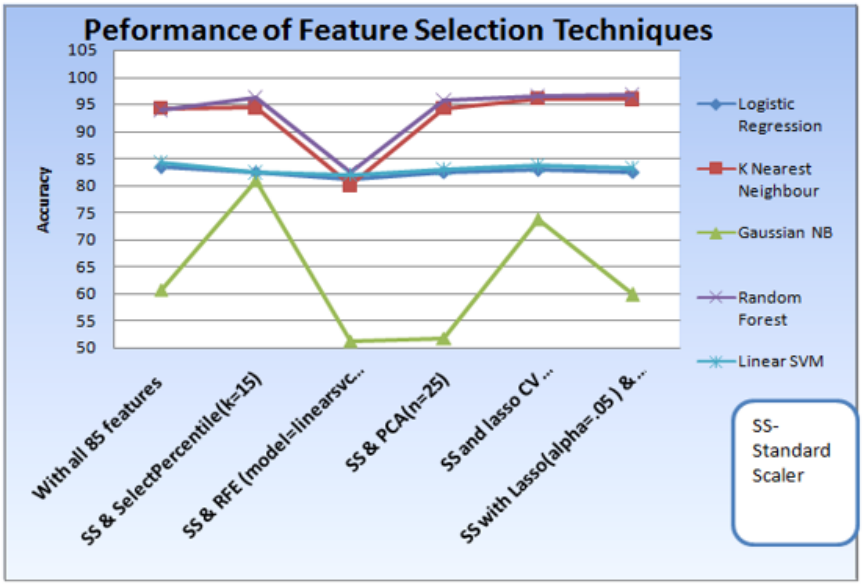

Fig. 4. Performance of machine learning models with different feature selection methods.

E. Discussion

We developed and used a Python-based software application for analyzing the network traffic dataset. The feature selection methods were chosen to evaluate all the three categories of feature selection/ elimination helped in generalizing the results. Results were obtained in a reasonable time frame. The finding of our experiments though ranked performances of the classifiers yet we cannot call any classifier 'the best' in every situation. Picking different machine learning algorithms alongside feature selection techniques altered the outcome which will be investigated in our future work.

In ordinary 'train_test_split' strategy', the machine learning models perform inadequately due to 'underfitting' or 'overfitting'. Overfitting refers to the learning of a machine learning model over the data and noise as well and adversely affects the performance of the models. Underfitting is another unwanted phenomenon when machine learning models fail to fit the training data well. There is every chance that the left-out data contain important patterns that induce the error by bias. Thus, we used K=10 Fold cross-validation for evaluation and comparison of our machine learning models and it indeed helps in avoiding these situations and results can be validated over the whole of the data.

F. Limitations

We evaluated our experiments with two metrics namely accuracy and number of features. More metrics as Precision, Recall, and F-measure can certainly present a more generalized view.

5. Conclusion

Feature selection is very vital from the perspective of the application of Machine learning algorithms. They boost the performance of these algorithms, reduces the complexity and time of execution. In other words, ML models become efficient with the application of feature selection. It removes irrelevant, redundant features without any considerable loss of information. Redundant features are those features that have do not have strong correlations with other features and thus their removal does not impact the outcome and the same applies to irrelevant features.

Detecting DDoS attacks at an early stage could spare the organizations from monitory, data and reputation loss. We evaluated feature selection methods one each from the entire three categories namely: Filter, Wrappers and Embedded methods. Our work has illustrated that feature selection has a direct impact on the outcome of machine learning-based classifiers for the detection of DDoS attacks. We validated the results using two metrics. The outcome of our experiments promptly concluded that KNN & Lasso CV with the SelectFromModel method and Random Forest alongside Lasso-RFE have performed better than the rest of the feature selection methods & classifiers. Further, we were successful in reducing the features from 85 to 12/ 15 without compromising on accuracy. Our results are more relevant as we used dataset having contemporary DDoS attacks. In fact, the accuracy of both the classifiers was improved with the feature selection. In the future, common features in multiple datasets will be evaluated as their impact on accuracy will be a fascinating study.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The data used to support the findings of this study are included in the article in table III. “The authors declare that there is no conflict of interest regarding the publication of this paper.”

References

[1] https://www.symantec.com/connect/blogs/iot-devices-being-increasingly-used-ddos-attacks

[2] S. KUMAR and K. M. CARLEY, *Understanding DDoS cyber-attacks using social media analytics*, 2016 IEEE Conference on Intelligence and Security Informatics (ISI), Tucson, AZ, pp. 231–236, 2016.

[3] P. FARINA, E. CAMBIASO, G. PAPALEO and M. AIELLO, *Understanding DDoS Attacks from Mobile Devices*, 2015 3rd International Conference on Future Internet of Things and Cloud, Rome, pp.614–619, 2015.

[4] Classification Accuracy, Available online,2019: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234

[5] NGUYEN H.T., PETROVIC S., FRANKE K., *A comparison of feature-selection methods for intrusion detection* LNCS 6258, 242–255, 2010.

[6] Herve NKIAMA, Syed Zainudeen Mohd SAID, Muhammad. SAIDU, *A Subset Feature Elimination Mechanism for Intrusion Detection System*, International Journal of Advanced Computer Science and Applications, **7**(4), 2016.

[7] O. OSANAIYE, K. R. DLODLO, M.D. LO, *Analysing Feature Selection and Classification Techniques for DDoS Detection in Cloud* Southern Africa Telecommunication Networks and Applications Conference (SATNAC) 2016, At Fancourt, George, Western Cape, South Africa.

[8] M. FATIH, S. BAYIR, *Examining the Impact of Feature Selection Methods on Text Classification*, International Journal of Advanced Computer Science and Applications **8**(12), January 2017.

[9] O. OSANAIYE, H. CAI, K.-K. R. CHOO, A. DEHGHANTANHA, Z. XU, M. DLODLO, *Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing*, EURASIP J. Wireless Commun. Netw., **2016**, pp. 130, Dec. 2016.

[10] A. PEKTA, T ACARMAN, *Effective Feature Selection for Botnet Detection based on detection based on network flow Analysis*, Conference: International Conference Automatics and Informatics'2017.

[11] KABIRI P, ZARGAR G (2009) Category-based selection of effective parameters for intrusion detection. Int J Com- put Sci Netw Secur (IJCSNS) **9**(9), 181–188.

[12] P. KALAPATAPU1, S. GOLI1, P. ARTHUM1, A. MALAPATI1, *A Study On Feature Selection And Classication Techniques Of Indian Music*, The 7th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2016), Procedia Computer Science **98**, 125–131, 2016.

[13] Iman SHARAFALDIN, Arash Habibi LASHKARI, and Ali A. GHORBANI, *Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization*, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018.

[14] Description of the CICIDS2017 dataset, available online, 2017, https://www.unb.ca/cic/datasets/ids-2017.html

[15] Feature Selection, Available online, 2018, https://en.wikipedia.org/wiki/Feature_selection

[16] Introduction to Feature Selection, Available online, 2018, https://machinelearningmastery.com/an-introduction-to-feature-selection

[17] Selectfrommodel: a feature selection method, Available online, 2017, http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectFromModel.html

[18] Selectpercentile: a feature selection method, Available online, 2017https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectPercentile.html

[19] https://en.wikipedia.org/wiki/Principal_component_analysis (accessed on 15-01-18)

[20] http://scikit-learn.org/stable/ (accessed on 20-02-18)

[21] Ayon Dey *Machine Learning Algorithms: A Review* International Journal of Computer Science and Information Technologies **7**(3), 2016, 1174–1179.

[22] https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/ (accessed on 19-02-18)

[23] Cross validation vs 'train_test_split' strategy https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation