Feature Selection in Machine Learning-Based IDS Performance

Jose Albeiro Montes Gil [1[0000-0002-7117-3051], Néstor Darío Duque Méndez [2[0000-0002-4608-281X], Gustavo Adolfo Isaza [3[0000-0002-1089-4605], Fabián Alberto Ramírez [4[0000-0002-6961-4604] and Jeferson Arango López [5[0000-0001-8072-9130]

[1] Universidad Nacional de Colombia, Colombia
[2] Universidad Nacional de Colombia, Colombia
[3] Universidad de Caldas, Colombia
[4] Universidad de Caldas, Colombia
[5]Universidad de Caldas, Colombia

Abstract. Computer security faces many challenges, including the detection of attacks generated by various intrusions. Intrusion detection systems (IDS) have different approaches, and those based on supervised learning have high capabilities to predict different types of attacks. As in other cases, machine learning algorithms supporting these tools benefit from dimensionality reduction. In this paper we present the results of applying different algorithms to obtain subsets of features that maintain good performance in classification tasks and apply the ensemble operations strategy to obtain the final features that are fed to the classifiers to determine the outputs. The results show that in the different classifiers all the metrics go down if the number of features is decreased in high degree, but that with the operation union of the subsets of the features obtained from the importance of the features, the number of attributes obtained is significantly lower and the good performance of the classifier is maintained.

Keywords: IDS, feature selection, dimensionality reduction.

1 Introduction.

In a world increasingly interconnected through computer networks and other devices, security in terms of privacy, reliability and availability is fundamental. The availability of systems is affected, among several situations, by various attacks from outside or inside the organization, including denial of service attacks, DoS, which reduce the availability of systems and even leave them completely inactive. Among the alternatives that have been implemented to deal with this situation are intrusion detection systems (IDS), which are systems that can determine whether security has been compromised at any time by permanently examining frame traffic. There are several approaches to IDS and several techniques that have been implemented to obtain a good performance of these, among them, artificial intelligence, and machine learning techniques, which in recent times have demonstrated the great capabilities that ensure systems with high performance achieving quite satisfactory metrics.

IDSs based on supervised learning have shown their great possibilities and capabilities to predict these situations with high performance Figs, but on many occasions,

with a high false positive rate. For training, a high volume of data is required and for the real-time execution phase, with already trained models, algorithms with acceptable computational cost must be implemented and high dimensionality negatively affects response times.

High dimensionality is recognized as one of the situations that can negatively influence machine learning algorithms. Reducing dimensionality under approaches such as relevant feature selection is a promising way, but it is not a trivial problem since the selected attributes change with different algorithms. For the case CICIDS-2017 data set, with 79 features originally and a class that reflects different types of attacks, it is very important to define a minimum number of features that allow a satisfactory classification or prediction, which forces to use computational methods to get those attributes with good performance while reducing the computational cost.

The work presented in this paper aims to run different algorithms that allow the selection of features and face the fact of obtaining non-homogeneous results in different cases, which maintains uncertainty. As a strategy, the subsets are integrated using the Union operator or the Intersection operator, seeking to determine the final features to be selected.

For the performance evaluation, different algorithms are applied to those that generate the subsets of features and are fed with the total of the attributes and with the features obtained with union and intersection. The results show that the best output is obtained, for the different metrics and algorithms, with the total features, but with the subset obtained with union the output drops very little in performance. The results of this work will be exploited in an IDS deployed in the cloud and for public use.

The rest of the paper is organized as follows: The next section is devoted to collect some concepts of interest for the paper, then some related work is discussed. Section 4 presents the methodology and Section 5 presents the validation and finally the Conclusions and Future work are included.

2 Theoretical Framework.

2.1 Denial of Service Attack (DoS).

A Denial of Service (DoS) attack is defined as the massive sending of requests to a service or equipment on the network, which generates a collapse in the receiver's resources, partially or totally affecting the availability of the service or equipment [1] .

2.2 Distributed Denial of Service Attack (DDoS).

Distributed Denial of Service attacks are requests coming from various sources, as shown in Fig 1. Unlike DoS attacks, in a DDoS attack common users can be infected under the concept of "zombie computers", which generates a greater number of requests, which fall on the receiver, causing response times to increase or the definitive collapse for an indeterminate period [1].

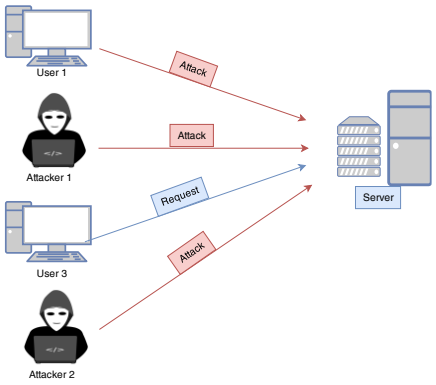


Fig 1: DDoS Attack.
Source: Own elaboration.

2.3 Dataset: CICIDS-2017.

The CICIDS-2017 Dataset was created by the University of New Brunswick in 2017 at the faculty of Computer Science. For 5 days a traffic analysis was performed on 25 users making use of HTTP, HTTPS, FTP, SSH and email protocols, which allowed the capture of 2'827,876 records. The analysis yielded 8 different types of attacks, including Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DdoS [2] [3].

The records were divided by days into files in CSV format, and DoS/DDoS attacks were used for this research. In total there are 1448366 rows and 79 columns, before preprocessing.

2.4 Intrusion Detection System.

An Intrusion Detection System is defined in [4] like a tool that can detect attacks both inside and outside the network, given its ability to monitor traffic and apply classification algorithms. An IDS can be classified into two categories: signature-based IDS and anomaly-based IDS. The operation of signature-based IDS relies on logs stored in databases, while anomaly-based IDS compares network patterns and behavior.

2.5 Feature Selection.

Attribute selection is a strategy often used in some research to reduce dimensionality by eliminating irrelevant or redundant attributes. As mentioned in [5], non-essential features negatively affect classifier performance and computational resource requirements.

Attribute selection is a strategy often used in some research to reduce dimensionality by eliminating irrelevant or redundant attributes. As mentioned in, non-essential features negatively affect classifier performance and computational resource requirements.

Since the techniques that allow attribute selection can vary in performance depending on the nature of the data, the general procedure for feature selection is defined in [5]. There, 4 stages are mentioned that can be applied to the CICIDS-2017 dataset, as shown in Fig 2.

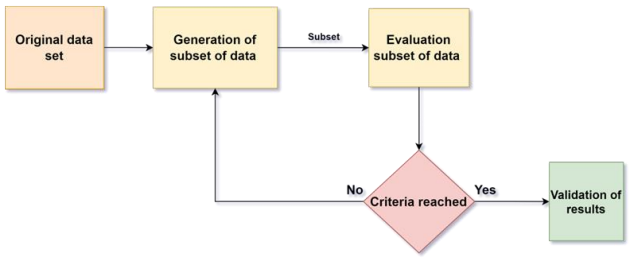


Fig 2: General Attribute Selection Procedure
Source: *From* [5].

There are several strategies and algorithms for feature selection, including Boruta, feature importance, principal component analysis (PCA), Chi-Square, among others.

2.6 Decision Tree (DT).

A Decision Tree is a non-parametric classification technique that uses rules based on the grouping of criteria [6]. Its graphical representation allows to clearly represent the events that are generated according to a decision [7].

2.7 Extreme Gradient Rise – XGBoost (XGB).

As mentioned in [8], XGBoost is a technique developed by Tianqi Chen which combines different decision trees that improve their performance with respect to those previously defined. In [9], the use of the XGBoost algorithm is highlighted as an approach successfully used in different models for intrusion detection.

As mentioned in, XGBoost is a technique developed by Tianqi Chen which combines different decision trees which improve their performance with respect to those previously defined. The use of the XGBoost algorithm is highlighted as an approach successfully used in different models for intrusion detection.

2.8 RandomForest (RF).

RandomForest is defined in [10] as a non-parametric classification technique, supported on trees that generate predictions based on random vectors. Its application is established in classification and regression tasks. In [10], the advantage of

RandomForest is highlighted, given its capacity for attribute selection and its resistance to overfitting.

2.9 K Nearby Neighbors – KNN.

The K Nearest Neighbors (KNN) algorithm is a nonparametric supervised learning technique that classifies by means of data proximity, since it performs its prediction by calculating the Euclidean distance between a value and the nearest records [11].

2.10 Multilayer Perceptron – MLP.

It is an artificial neural network which is composed of several hidden layers, which in turn constitute the input of another layer of neurons. An MLP neural network has a set of input neurons, a layer of hidden neurons and the neurons that constitute the prediction [12].

3 Related Research.

The following are some papers presenting strategies for feature selection in data sets like the one exploited in this work.

In [2] the authors applied Boruta and Permutation Importance for Feature Selection of the CICIDS-2017 dataset. Different Supervised Learning techniques were applied to analyze the accuracy of each of the classifiers, among which RandomForest and Decision Trees are highlighted. The authors mention the importance of continuing to perform analyses with other types of classifiers and feature selectors.

In [13] a proposal was presented based on Feature Selection using Filter and Wrapper based techniques on the NSL-KDD dataset. Subsequently, the authors built an Intrusion Detection System using Machine Learning techniques with the features obtained by Chi-Square and Correlation. The results obtained reflect a higher accuracy of the Artificial Neural Network compared to the Support Vector Machine. The authors emphasize the importance of continuing research in the field of intrusions, given the false positive rates that are still occurring.

In [14] the CSE-CIC-IDS2018 dataset was used for the application of Machine Learning and Feature Selection techniques with the implementation of Chi-Square and Spearman's Correlation Coefficient. The model proposed by the authors was validated using 7 different classifiers, however, as future work, the need to advance studies under other hyperparameters and different types of Artificial Neural Networks is proposed.

In [15] a scheme for an IDS divided in 2 stages was built. Stage 1 consisted of data preprocessing and Feature Selection, and stage 2, the use of the classifier supported by Light Gradient Boosting Machine. Comparisons were made with other techniques such as Support Vector Machines, RandomForest, Convolutional Neural Networks, Multilayer Perceptron, among others.

In [16] the authors used the NSL-KDD dataset with random features to reduce the number of attributes, training times and model complexity. They implemented 8

Machine Learning techniques, which showed a good behavior in DoS attacks, while in U2R attacks their performance was inferior. As future work, the importance of implementing different methods for Feature Selection to reduce the computational complexity of IDS proposals is mentioned.

In [17] Chi-Square was applied as a feature selection technique for the UNSW-NB15 dataset. The performance of the 5 Machine Learning classification algorithms was analyzed in terms of accuracy, false positive rate, F1 score and mean square error. The analysis was performed under two proposals; the first one considered the dataset considering the selection technique, while the second one, did not consider the feature selection technique. This paper concludes that the Random Forest classifier was the best performing classifier compared to Naive Bayes, Logistic Regression and KNN.

In [18] the authors conducted a review of papers around the field of IDS, which were implemented under Machine Learning and Deep Learning techniques. Twenty-eight papers were analyzed, defining which is the most analyzed dataset and the most frequent learning technique. As a conclusion, the authors highlight the importance of performing validations with different data sets, given that most of the research is based on the KDD Cup'99 and NSL-KDD data set. Additionally, it is recommended to find mechanisms through feature selection to reduce the dimensionality of the problems and increase the performance of the classification models.

In [19] , a comparison is made between normalization and feature selection using the UNSW-NB15 dataset, with the aim of determining which aspect is more relevant when analyzing the performance of a model. In this paper, the authors conclude that the evaluated models performed better when trained with recent datasets like UNSW-NB15 compared to older ones, such as NSL-KDD.

In [20] a critical review of other studies was carried out in terms of the most analyzed data sets, feature selection techniques and evaluation of the proposals in terms of parameters such as complexity, correlation, and performance. The importance of including techniques that allow the selection of features that can make significant contributions, considering the number of variables to be analyzed, is highlighted.

The review of research articles in the field of intrusion detection conducted in [21] includes 29 papers that use different mechanisms for Feature Selection and Classification algorithms supported by Machine Learning. It also highlights the popularity of Artificial Neural Networks in their different classifications.

While feature reduction can generate a reduction in the complexity of a model, in [22] a higher performance in terms of accuracy is observed when the number of attributes increases. In [23] the performance of some algorithms applying feature selection was analyzed, from there it is concluded that the accuracy of the models can be increased in some cases as the number of features increases.

The literature review conducted allows observing that feature selection in computer security related datasets is a field of interest and can bring improvements to the performance of classification models, however, a unique feature subset that delivers the best results in different classifiers is not determined.

Table 1 below presents a summary of the focus of each of the papers that make up the state-of-the-art review.

The following convention is used:

- Meets criterion (x).
- Does not meet criterion (-).

Table 1: Summary of research approach.

Research	Define Importance of each attribute	Dataset CICIDS 2017	Presents subassembly integration strategy	Comparison between experiment using all attributes and using attribute selection
[2]	**-**	**X**	**-**	**X**
[9]	**-**	**-**	**-**	
[10]	**X**	**-**	**-**	**X**
[11]	**-**	**-**	**-**	**-**
[13]	**-**	**-**	**-**	**-**
[24]	**-**	**-**	**-**	**-**
[25]	**-**	**-**	**-**	**X**
[26]	**-**	**-**	**-**	**-**

Source: Own elaboration.

As shown in Table 1, the related works do not reflect a tendency to determine the most important characteristics in the generation of DoS/DDoS attacks, in addition to the absence of proposals that seek to evaluate the results from subsets of data defined by different classifiers. None of the works present the different characteristics selected by different methods used and a strategy to determine the final subset.

4 Metodology

This work was developed following a strategy divided into 3 stages, as shown in Fig 3. With the proposed segmentation by stages, it is expected that it will be easier for future interested parties to replicate the present research. The stages defined in the present methodology are supported by what is mentioned in [2] and [5] for the definition of the criteria in the preprocessing and the stop condition to perform the validation of the results obtained in the feature selection.

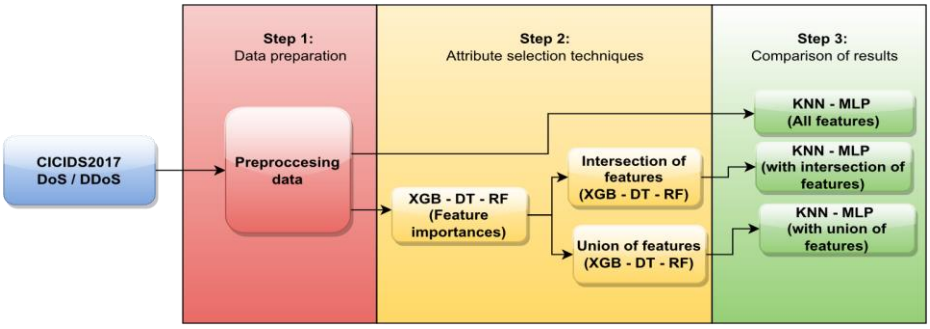


Fig 3: Methodology by Stages
Source: Own elaboration

4.1 Stage 1- Data preparation.

Data preparation is the stage in which the aim is to manipulate the records in such a way that they are useful for further analysis. The objective is to eliminate noise, subtract outliers, complete missing records, among others [27] . Based on the CICIDS-2017 data set, records with trends to infinity, non-numerical (Not a Number), blank spaces, a repeated column, replacing attribute names with values from 0 to 78 were eliminated, as shown in Table 2.

Table 2: CICIDS2017 Columns renamed.

Number	Column CICIDS2017	Number	Column CICIDS2017	Number	Column CICIDS2017
1	dst_port	27	bwd_iat_mean	53	pkt_size_avg
2	flow_duration	28	bwd_iat_std	54	fwd_seg_size_avg
3	tot_fwd_pkts	29	bwd_iat_max		
4	tot_bwd_pkts	30	bwd_iat_min	55	bwd_seg_size_avg
5	totlen_fwd_pkts	31	fwd_psh_flags		
		32	bwd_psh_flags	56	fwd_byts_b_avg
6	totlen_bwd_pkts	33	fwd_urg_flags		
		34	bwd_urg_flags	57	fwd_pkts_b_avg
7	fwd_pkt_len_max	35	fwd_header_len		
		36	bwd_header_len	58	fwd_blk_rate_avg
8	fwd_pkt_len_min	37	fwd_pkts_s		
		38	bwd_pkts_s	59	bwd_byts_b_avg
9	fwd_pkt_len_mean	39	pkt_len_min		
		40	pkt_len_max	60	bwd_pkts_b_avg
10	fwd_pkt_len_std	41	pkt_len_mean		
		42	pkt_len_std	61	bwd_blk_rate_avg
11	bwd_pkt_len_max	43	pkt_len_var		
		44	fin_flag_cnt	62	sub-flow_fwd_pkts
12	bwd_pkt_len_min	45	syn_flag_cnt		
		46	rst_flag_cnt	63	sub-flow_fwd_byts
13	bwd_pkt_len_mean	47	psh_flag_cnt		
		48	ack_flag_cnt	64	sub-flow_bwd_pkts
14	bwd_pkt_len_std	49	urg_flag_cnt		
		50	cwe_flag_count		
15	flow_byts_s	51	ece_flag_cnt		
16	flow_pkts_s	52	down_up_ratio		

17	flow_iat_mean	65	sub-flow_bwd_byts
18	flow_iat_std	66	init_fwd_win_byts
19	flow_iat_max	67	init_bwd_win_byts
20	flow_iat_min	68	fwd_act_data_pkts
21	fwd_iat_tot	69	fwd_seg_size_min
22	fwd_iat_mean	70	active_mean
23	fwd_iat_std	71	active_std
24	fwd_iat_max	72	active_max
25	fwd_iat_min	73	active_min
26	bwd_iat_tot	74	idle_mean
		75	idle_std
		76	idle_max
		77	idle_min
		78	Label

Source: Own elaboration.

Additionally, the values were normalized. Table 3 shows the amount of data before and after the preprocessing stage.

Table 3: Variations in data with preprocessing.

Item	**CICIDS-2017 (DoS – DDoS) without prepro-cessing**	**CICIDS-2017 (DoS – DDoS) with preprocessing**
Rows	1448355	1447279
Columns	79	78

Source: Own elaboration

Table 4 shows the coding process to which the classes were subjected, where the outputs are represented by numerical values between 0 and 5.

Table 4: Class Categorization.

Label	**Class**	**Records**
0	Benigno	1067540
1	DDoS	128027
2	DoS GoldenEye	10293
3	DoS Hulk	230124
4	DoS Slowhttptest	5499
5	DoS Slowloris	5796
	Total	**1447279**

Source: Own elaboration.

4.2 Stage 2- Application of techniques for attribute selection.

Fig 4 shows that stage 2 consisted of identifying the percentage of importance of each of the attributes in the dataset. To determine the importance of the features, the XGB, RF and DT classifiers were defined using the well-known free software library for machine learning Scikit-Learn.

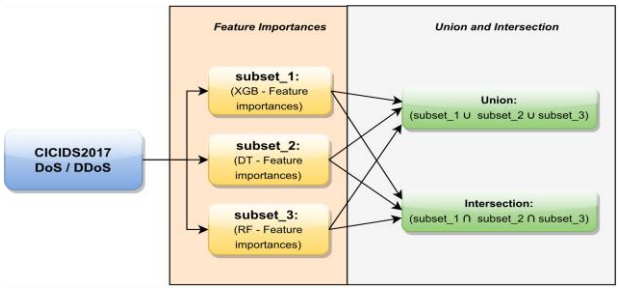


Fig 4: Attribute selection stage.
Source: Own elaboration.

To avoid overfitting, the data set was divided into 75% of the records for the training process and 25% for the tests. These values were defined from experience in previous work and guidelines in the literature.

To determine the importance of each of the features, the attribute "feature_importances_" of the Scikit-Learn library is used, as recommended in [28], [29] y [30].
Fig 5 shows the performance of the classifiers in terms of precision, F1 score, accuracy and sensitivity. It is possible to observe that the 3 techniques obtained significant metrics, when the model was validated with 25% of the data.

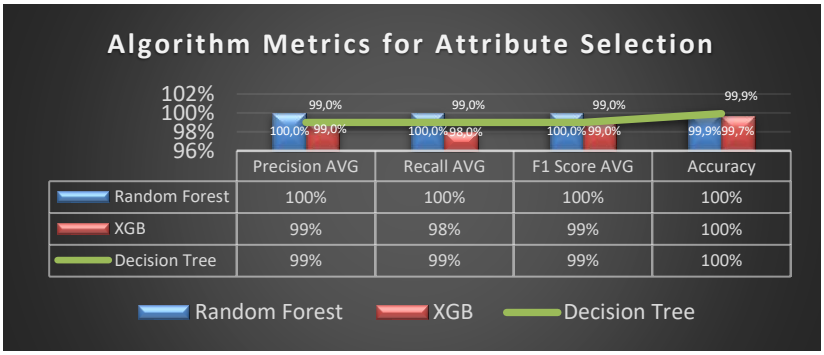


Fig 5: Classifier Performance - Attribute Importance.
Source: Own elaboration.

The chosen classifiers are run on the preprocessed data set and give a weighting of the weight of each attribute in obtaining the classifier results. Tables 5, 6 and 7 show the percentage of importance of each attribute according to each classification technique. Since "feature_importances" determines in percentage the importance of each attribute, the cumulative sum of the attributes was considered for the creation of each subset until the percentage of importance reached 80%.

Table 5: Feature Importance XGB.

Feature	% Importance	% of Importance (cumulative)
3	16,00%	16%
35	8,67%	25%
36	6,42%	31%
38	6,35%	37%
4	6,16%	44%
25	5,26%	49%
0	3,85%	53%
7	3,29%	56%
42	3,20%	59%
6	2,93%	62%
67	2,89%	65%
75	2,79%	68%
11	2,48%	70%
22	2,27%	73%
24	2,18%	75%
65	2,05%	77%
14	1,78%	79%
23	1,76%	80%

Source: Own elaboration.

Table 6: Feature Importance DT.

Feature	% Importance	% of Importance (cumulative)
75	22,50997%	22,50997%
42	17,87567%	40,38564%
0	13,88353%	54,26917%
36	13,20165%	67,47083%
71	8,34455%	75,81537%
6	4,38788%	80,20326%

Source: Own elaboration.

Table 7: Feature Importance RF.

Feature	% Importance	% of Importance (cumulative)
36	5,366%	5,366%
75	4,782%	10,148%
73	4,638%	14,786%
76	4,436%	19,222%
0	4,382%	23,604%
1	4,138%	27,742%
22	3,864%	31,606%
42	3,615%	35,221%
65	3,533%	38,754%
15	3,461%	42,215%
23	3,299%	45,514%
4	3,064%	48,578%
13	2,761%	51,339%
14	2,584%	53,923%
21	2,352%	56,275%
20	2,220%	58,495%
62	1,957%	60,452%
17	1,779%	62,231%
16	1,758%	63,989%
18	1,747%	65,735%
34	1,727%	67,462%
69	1,650%	69,112%
5	1,632%	70,744%
71	1,611%	72,355%
6	1,584%	73,940%
10	1,479%	75,419%
72	1,401%	76,819%
27	1,389%	78,208%
25	1,373%	79,581%
9	1,363%	80,943%

Source: Own elaboration

Table 8 presents the characteristics obtained with each of the classifiers.

To compare the performance of the model using all the features and attribute selection, it was defined to perform the intersection and union between the 3 subsets of features obtained by each classifier, obtaining 5 features in the intersection (Table 9) and 37 with the union (Table 10).

Table 8: Attribute subsets for each classifier.

Subset	Features
subconjunto_RF	**36, 75, 73, 76, 0, 1, 22, 42, 65, 15, 23, 4, 13, 14, 21, 20, 62, 17, 16, 18, 34, 69, 5, 71, 6,10, 72, 27, 25, 9**
subconjunto_XGB	**3, 35, 36, 38, 4, 25, 0, 7, 42, 6, 67, 75, 11, 22, 24, 65, 14, 23**
subconjunto_DT	**75, 42, 0, 36, 71, 6**

Source: Own elaboration.

Table 9: Intersection of attributes.

Subset	Features
subset_intersección	0, 36, 6, 42, 75

Source: Own elaboration

Table 10: Union of attributes.

Subset	Features
subset_union	0, 1, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 27, 34, 35, 36, 38, 42, 62, 65, 67, 69, 71, 72, 73, 75, 76

Source: Own elaboration.

4.3 Stage 3 - Comparison of Results.

This stage is defined with the objective of comparing the metrics of the KNN classifiers and the NN-MLP multilayer perceptron neural network with the subsets defined in the feature selection stage, as shown in Fig 6.

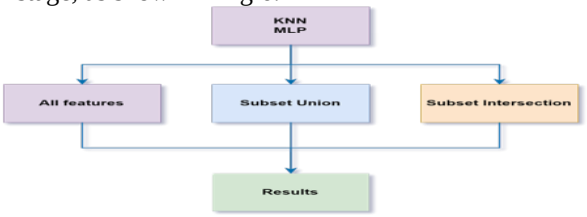


Fig 6: Application of subsets to KNN and MLP.
Source: Own elaboration.

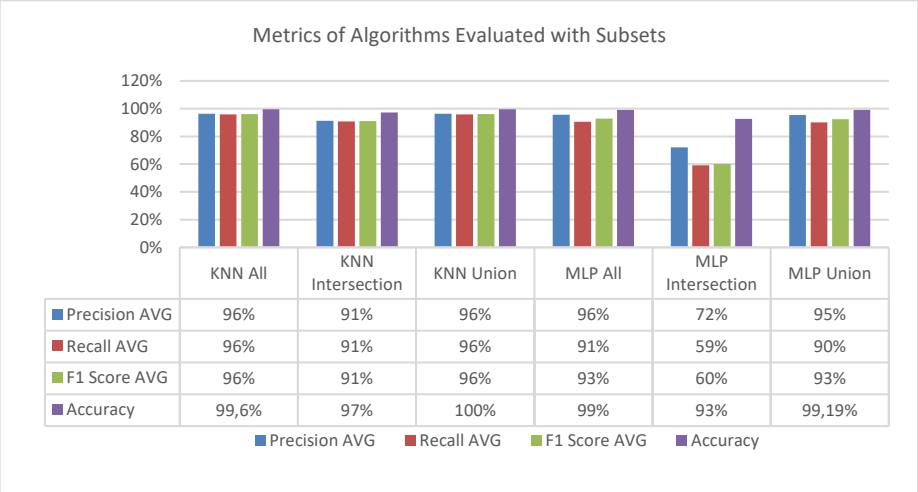
Stage 3 ends with the analysis of the KNN and MLP classifier metrics trained on the intersection, union and full-featured dataset subsets.

5 Validation With Feature Subset.

All calculations were performed using the Google Colab environment under Python 3.7.15. The code is available at https://acortar.link/c9PTFm

5.1 Evaluation of KNN and MLP Models Using Subsets.

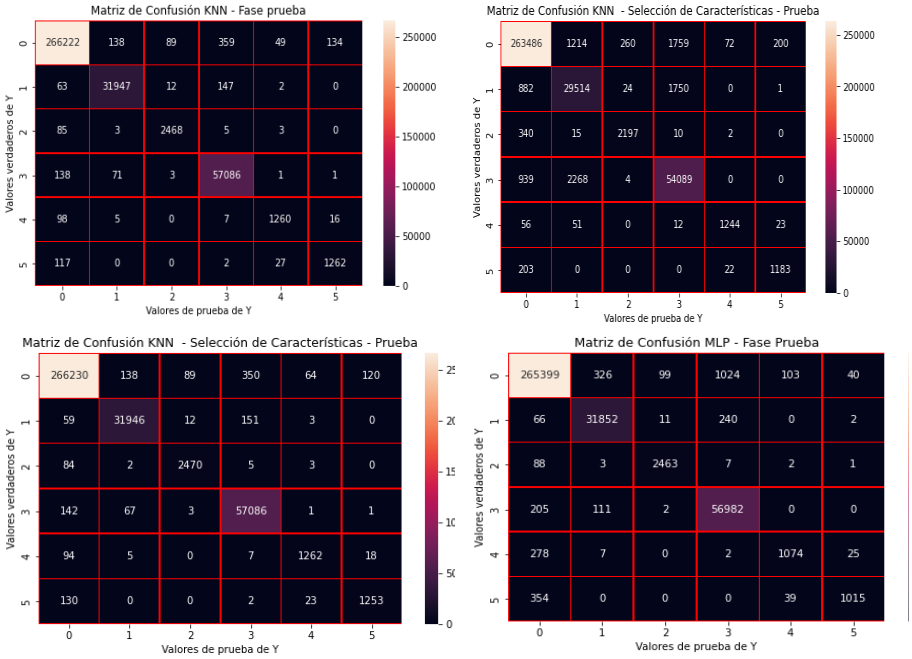
To validate the relevance of the features selected in the previous phase they build the models with KNN and MLP. Fig 7 shows the summary of the metrics in terms of precision, accuracy, F1 score and sensitivity.



	KNN All	KNN Intersection	KNN Union	MLP All	MLP Intersection	MLP Union
Precision AVG	96%	91%	96%	96%	72%	95%
Recall AVG	96%	91%	96%	91%	59%	90%
F1 Score AVG	96%	91%	96%	93%	60%	93%
Accuracy	99,6%	97%	100%	99%	93%	99,19%

Fig 7: Evaluation of Metrics with Subsets and all Attributes.
Source: Own elaboration.

As shown in Fig 7, the KNN model yielded significant metrics in the 3 moments in which it was evaluated (see Figs 8, 9 and 10); however, its performance improved as the number of attributes increased. On the other hand, the MLP model did not perform as well (see Figs 11, 12 and 13) and as the number of attributes decreased, the quality of the classification decreased. The KNN model was able to group the different DoS/DDoS attacks and the traffic identified as benign, as can be seen in the different confusion matrices (Figs 8, 9 and 10).



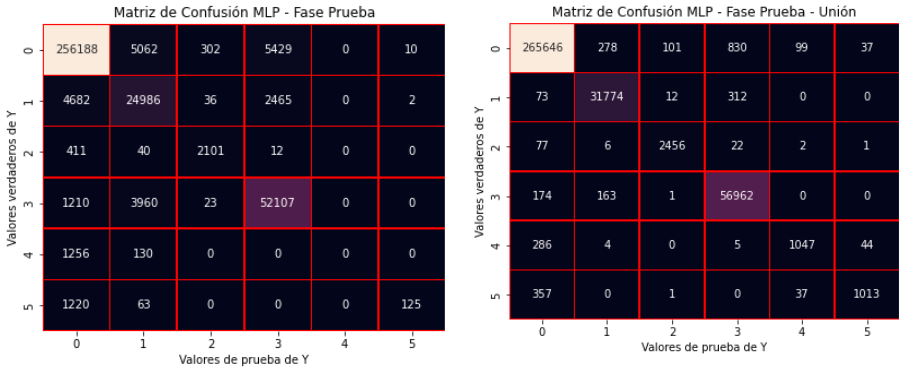


Fig 8: Confusion matrices**.**
Source: Own elaboration.

6 Conclusions and Future Work.

In this paper, an experimental study was conducted on the classification of different types of DDoS/DoS attacks defined in the CICIDS2017 data set. Given the large number of attributes and instances in the dataset, a proposal was made to evaluate different classification techniques using all attributes and analyzing the metrics when a reduction to the dimensionality of the data is performed. The attribute selection algorithm yielded different subsets of features with different classifiers. In order to determine the behavior of other classifiers with the obtained features, they were run with the total features, intersection and union.

The results show that when having few features (those obtained with the intersection) the performance is reduced. But the union, which yields 37 attributes, has a performance very close to the one obtained with the total of features (78), in the 2 classifiers with which the validation was performed and for all metrics.

This result, which is not conclusive, is a contribution that can improve the effectiveness of intrusion detection systems aimed at improving computer security by inspecting significant features in the classification process.

Future work is expected to validate the results obtained with other data sets, including those obtained with actual traffic. Tests will also be performed with variations of the models with different hyperparameters and other classification techniques, as well as including data obtained from DDoS attacks generated with new signatures.

As a work in progress, the results of this work will be leveraged in an IDS deployed in the cloud and for public use.

References

1. Hadeel S. Obaid, “Denial of Service Attacks: Tools and Categories,” International Journal of Engineering Research and, vol. V9, no. 03, pp. 631–636, 2020, doi: 10.17577/ijertv9is030289.
2. E. M. Ortiz Martínez, P. Arguijo, A. Hiram Vázquez López, R. Ángel, and M. Armenta, “Selección de características con método wrapper para un sistema de detección de intruso: caso CICIDS-2017 Feature Selection with a Wrapper Method for Intrusion Detection System: Case CICIDS-2017,” 2020.
3. University of New Brunswick, “Intrusion Detection Evaluation Dataset (CIC-IDS2017).” Accessed: Sep. 28, 2021. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html
4. S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, and H. Karimipour, “Cyber intrusion detection by combined feature selection algorithm,” Journal of Information Security and Applications, vol. 44, pp. 80–88, Feb. 2019, doi: 10.1016/j.jisa.2018.11.007.
5. V. Kumar and M. Sonajharia, “Feature Selection: A literature Review,” The Smart Computing Review, vol. 4, no. 3, Jun. 2014, doi: 10.6029/smartcr.2014.03.007.
6. B. Mahbooba, M. Timilsina, R. Sahal, and M. Serrano, “Explainable Artificial Intelligence (XAI) to Enhance Trust Management in Intrusion Detection Systems Using Decision Tree Model,” Complexity, vol. 2021, 2021, doi: 10.1155/2021/6634811.
7. C. Layme Fernández et al., “Application of decision trees in the identification of fraudulent websites,” Revista Innovación y Software, vol. 3, no. 1, 2022.
8. Y. Song, H. Li, P. Xu, and D. Liu, “A Method of Intrusion Detection Based on WOA-XGBoost Algorithm,” Discrete Dyn Nat Soc, vol. 2022, pp. 1–9, Feb. 2022, doi: 10.1155/2022/5245622.
9. P. Bedi, N. Gupta, and V. Jindal, “I-SiamIDS: An Improved Siam-IDS for handling class imbalance in Network-based Intrusion Detection Systems,” 2021.
10. M. Choubisa, R. Doshi, N. Khatri, and K. K. Hiran, “A Simple and Robust Approach of Random Forest for Intrusion Detection System in Cyber Security,” in 2022 International Conference on IoT and Blockchain Technology, ICIBT 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICIBT52874.2022.9807766.
11. H. González, G. Santos, F. Campos, and C. Morell Pérez, “Evaluación del algoritmo KNN-SP para problemas de predicción con salidas compuestas Evaluation of KNN-SP algorithm for multi-target prediction problems,” Revista Cubana de Ciencias Informáticas, vol. 10, no. 3, 2016, [Online]. Available: http://rcci.uci.cuPág.119-129
12. Y. Yigit, B. Bal, A. Karameseoglu, T. Q. Duong, and B. Canberk, “Digital Twin-Enabled Intelligent DDoS Detection Mechanism for Autonomous Core Networks,” IEEE Communications Standards Magazine, vol. 6, no. 3, pp. 38–44, Sep. 2022, doi: 10.1109/MCOMSTD.0001.2100022.
13. Kazi Abu, Y. J. Billal Mohammed, and R. Md. Mahbubur, Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection. IEEE, 2019.
14. S. F. Qusyairi Ridho and R. Kalamullah, “Implementation of Ensemble Learning and Feature Selection for Performance Improvements in Anomaly-Based Intrusion Detection Systems,” 2020.
15. Y. Hua, An Efficient Traffic Classification Scheme Using Embedded Feature Selection and LightGBM. 2020.
16 . A. Iram, A. Zahrah, M. Faheem, and B. Alwi M, A Machine Learning Approach for Intrusion Detection System on NSL-KDD Dataset. 2020.

17 . G. Kocher and G. Kumar, “Analysis of Machine Learning Algorithms with Feature Selection for Intrusion Detection using UNSW-NB15 Dataset,” International Journal of Network Security & Its Applications, vol. 13, no. 1, pp. 21–31, Jan. 2021, doi: 10.5121/ijnsa.2021.13102.
18. Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, “Network intrusion detection system: A systematic study of machine learning and deep learning approaches,” Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, Jan. 2021, doi: 10.1002/ett.4150.
19. M. Albarka Umar, Z. Chen, K. Shuaib, and Y. Liu, “Effects of Feature Selection and Normalization on Network Intrusion Detection,” COMMUNICATION, NETWORKING AND BROADCAST TECHNOLOGIES, pp. 1–27, Jan. 2024, doi: 10.36227/techrxiv.12480425.v3.
20. M. di Mauro, G. Galatro, G. Fortino, and A. Liotta, “Supervised feature selection techniques in network intrusion detection: A critical review,” Eng Appl Artif Intell, vol. 101, May 2021, doi: 10.1016/j.engappai.2021.104216.
21. C. Kalimuthan and J. Arokia Renjit, “Review on intrusion detection using feature selection with machine learning techniques,” in Materials Today: Proceedings, Elsevier Ltd, 2020, pp. 3794–3802. doi: 10.1016/j.matpr.2020.06.218.
22. A. Nazir and R. A. Khan, “A novel combinatorial optimization based feature selection method for network intrusion detection,” Comput Secur, vol. 102, Mar. 2021, doi: 10.1016/j.cose.2020.102164.
23. R. A. R. Mahmood, A. H. Abdi, and M. Hussin, “Performance evaluation of intrusion detection system using selected features and machine learning classifiers,” Baghdad Science Journal, vol. 18, pp. 884–898, Jun. 2021, doi: 10.21123/bsj.2021.18.2(Suppl.).0884.
24. Z. Liu, X. Yin, and Y. Hu, “CPSS LR-DDoS Detection and Defense in Edge Computing Utilizing DCNN Q-Learning,” IEEE Access, vol. 8, no. 3, pp. 42120–42130, 2020, doi: 10.1109/ACCESS.2020.2976706.
25. S. Xiao and W. Tong, “Prediction of User Consumption Behavior Data Based on the Combined Model of TF-IDF and Logistic Regression,” in Journal of Physics: Conference Series, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1757/1/012089.
26. P. Kanimozhi and T. Aruldoss Albert Victoire, “Oppositional tunicate fuzzy C-means algorithm and logistic regression for intrusion detection on cloud,” Concurr Comput, vol. 34, no. 4, Feb. 2022, doi: 10.1002/cpe.6624.
27. P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, “New data preprocessing trends based on ensemble of multiple preprocessing techniques,” TrAC - Trends in Analytical Chemistry, vol. 132. Elsevier B.V., Nov. 01, 2020. doi: 10.1016/j.trac.2020.116045.
28. J. Wang, X. Chang, Y. Wang, R. J. Rodríguez, and J. Zhang, “LSGAN-AT: enhancing malware detector robustness against adversarial examples,” Cybersecurity, vol. 4, no. 1, Dec. 2021, doi: 10.1186/s42400-021-00102-9.
29.J. L. Leevy, J. Hancock, R. Zuech, and T. M. Khoshgoftaar, “Detecting cybersecurity attacks across different network features and learners,” J Big Data, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00426-w.
30. N. v. Sharma and N. S. Yadav, “An optimal intrusion detection system using recursive feature elimination and ensemble of classifiers,” Microprocess Microsyst, vol. 85, Sep. 2021, doi: 10.1016/j.micpro.2021.104293.