Bayesian Classifiers in Intrusion Detection Systems

Mardini-Bovea Johan, De-La-Hoz-Franco Emiro, Molina-Estren Diego, Paola Ariza-Colpas, Ortíz Andrés, Ortega Julio, César Cárdenas, Carlos Collazos-Morales

Bayesian Classifiers in Intrusion Detection Systems

Mardini-Bovea Johan[1], De-La-Hoz-Franco Emiro[2], Molina-Estren Diego[3], Paola Ariza-Colpas[3], Ortíz Andrés[4], Ortega Julio[5], César A. Cárdenas R[6], and Carlos Collazos-Morales[6]

[1] Universidad de la Costa and Universidad del Atlántico, Barranquilla, Colombia
[2] Computer Science and Electronics Department, Research group Software Engineering and Networks, Universidad de la Costa, Barranquilla, Colombia
[3] Computer Science and Electronics Department, Research group Software Engineering and Networks, Universidad de la Costa, Barranquilla, Colombia
[4] Communications Engineering Department, University of Málaga, Málaga, Spain
[5] Computer Architecture and Technology Department, CITIC, University of Granada, Granada, Spain
[6] Vicerrectoría de Investigaciones, Universidad Manuela Beltrán, Bogotá, Colombia
edelahoz@cuc.edu.co

Abstract. To be able to identify computer attacks, detection systems that are based on faults are not dependent on data base upgrades unlike the ones based on misuse. The first type of systems mentioned generate a knowledge pattern from which the usual and unusual traffic is distinguished. Within computer networks, different classification traffic techniques have been implemented in intruder detection systems based on abnormalities. These try to improve the measurement that assess the performance quality of classifiers and reduce computational cost. In this research work, a comparative analysis of the obtained results is carried out after implementing different selection techniques such as Info. Gain, Gain ratio and Relief as well as Bayesian (Naïve Bayes and Bayesians Networks). Hence, 97.6 % of right answers were got with 13 features. Likewise, through the implementation of both load balanced methods and attributes normalization and choice, it was also possible to diminish the number of features used in the ID classification process. Also, a reduced computational expense was achieved.

Keywords: Naïve Bayes · Bayesian Networks · Feature Selection · IDS

1 Introduction

Cyber-attacks keep being a big problem in the current productive context. they can lead to the loss of sensitive information employed to make decisions within organizations. Thus, the necessity to develop tools to mitigate vulnerabilities in computer environments comes up. Several systems protect from malware data have emerged. However, when the database is not updated frequently, these systems may not be fully effective. As new attacks are created, inadequate management of vulnerabilities may generate catastrophic situations. Therefore, the detection of fraudulent actions has become one of the research priorities of information security. For this reason, algorithms have been integrated to many Intrusion Detection Systems - IDS based on data mining techniques for identification of anomalous traffic. [13], [17], [18], [15], [14] and [12]. The CERT (Information Security Incident Response Center) [13], has analyzed and classified in its database over 10,000 viruses. The viruses identified show that some remain over time and others evolve and adapt to new operating systems expanding their actions. It is feasible to provide users with more optimal tools for protection of different threats that may arise. INTECO has documented an important number of viruses that affect the operating systems of mobile devices, PCs and servers.

Some simulation environments have been developed which allow the design and implementation of new IDS Intrusion Detection Systems based on intelligent techniques [18] and [15]. The proposed models are evaluated through different experimental works in which quality measures are analyzed to be then implemented in productive environments, [3], [9]. Some of the techniques considered are Naïve Bayes, J48 and PART classifiers and Chi Square selection techniques and Consistency [16], tthe IBK classifier and the combination of Symmetric and Gain ratio selection techniques [20], assembled vector support classifiers and non-linear projection techniques [8], Bayesian authorizing

maps [4], Hybridization of statistical techniques and SOM performing feature selection with PCA + FDR [20], a wrapper-based method, applied using a multi-objective approach and using the GHSOM classifier [20]. This work focuses on the bi-class classification processes because of their relevance in real application situations where possible attacks are sought. In addition, it would be required to take corrective actions against the anomalous behavior that has been identified. Selection techniques have been applied based on information filtering: Info.Gain [3], Gain ratio [1] and Relief [11]. The main purpose is to identify the attributes that contribute the most to the classification process. Then, an appropriate selection technique is identified and applied [7], [10]. A comparative analysis of the quality metrics generated is made.

2 Pre-Processing

The simulation process used to validate the detection rates of a classifier implies the execution of a series of phases: pre-processing, selection, training/classification and evaluation of the performance of the classifier. The pre-processing phase involves the use of a data set from which the data to be analyzed comes from. In this type of research, the DARPA NSL-KDD data set has been used as it is widely supported by the scientific community that evaluates related studies. According to [14], there are some improvements that NSL-KDD has over its predecessors. The fact that it does not include redundant records in the data collection for training. There are no duplicated records in the data collections proposed for the tests. The number of records selected from each group of difficulty level is inversely proportional to the percentage of records in the original set of KDD data. Also, the number of records in the data collections for training and testing is reasonable. In the different simulation contexts described later, the NSL-KDD data set created from the KDD'99 [12] was used. The size of the NSL-KDD is smaller than the one of KDD'99 because the records of redundant connections have been eliminated. The NSL-KDD is made up of the KDDTrain+, KDDTrain+20Percent, KDDTest+ and KDDTest-21 files which are in both TXT and ARFF formats.

To be able to adjust the NSL-KDD set, techniques such as pre-processing, load balancing and normalization were applied. The load balancing is intended to level the number of normal connections and the number of attacks to avoid bias. Classifiers that are trained with unbalanced data sets, tend to classify data instances as part of the main class and ignoring the low representation of the minority class. Table 1 shows the amount of both the normal connections and the connections that represent attacks. They are contained in the NSL-KDD. 53.46% of the connections are normal exceeding by 6.92% the connections corresponding to the attacks. Thus, a load balancing technique called Synthetic Minority Oversampling Technique - SMOTE [13] was implemented. According to [19], this technique is responsible for adding random information to the training process of the data set generating new data instances. In this research work, SMOTE gives new instances of the "attack" data class by 14.86% of the current ones in the training data set NSL-KDD. Each new instance is computed from the average of the five closest neighbors and with a seed set to one.

Table 1: Connection Distribution for NSL-KDD Train

Training		
Connections	Qty	%
Normal	67.343	53.46
Attacks	58.630	46.54
Total	**125.973**	**100.00**

Regarding standardization, 41 features of the NSL-KDD data set are used in the different classification techniques. Therefore, the variables scale is very important to determine the topological organization of the structures used by these techniques. If the range of values of a variable is greater than the others, it will probably dominate the organization of the classifier structure. Normalization

prevents one characteristic from contributing more than another to the measurement of distances. In [6] six standardization methods are presented and have been evaluated in this proposal. According to the results acquired, it has been demonstrated that the technique with the best performance is the one called normalization at zero mean and unit variance. The continuous variables are normalized with zero mean and unit variance by using equation 1. On the other hand, all the variables are scaled at the interval [0...1]. The symbolic features and the binary ones are not normalized. The normalization technique employed is a simple linear transformation as shown in the following equation:

$$\hat{x} = \frac{x - \underline{x}}{\sigma} \tag{1}$$

Where x and $\underline{x}$ are the mean and the standard deviation of the variable x. This is equivalent to express the x variable as the distance between the number of deviations and its mean. After refining the data set through the pre-processing techniques mentioned previously, the different features selection methods are evaluated. The purpose is to reduce the complexity of the appraising process which will be then executed by the classifier.

3 Feature Selection Techniques

The feature selection phase is essential for an efficient analysis of the data contained in the data set since this usually contains information that adds noise to the generation process of the model. Because of this issue, there is some degradation of the quality of the patterns to be detected. The redundant variables and the irrelevant ones make it difficult to get significant patterns from the data. In In [16],it is stated that the ability to use a feature selection is essential to perform an effective analysis because the data contains information that is not necessary for the generation of the model. It is affirmed in [20] that the features selection allows to reduce the entries of the data to an appropriate size for processing and analysis. Therefore, attributes or features must be selected or discarded depending on their usefulness for the analysis. Every selection process of attributes has a starting point, which can be the complete set of attributes, the empty set or any intermediate state. After the first subset is evaluated, other subsets will be examined based on a search direction that can be forward, backward, random or any variation of the above. The process will finish when the entire space is covered or when a stop condition is fulfilled depending on the search strategy followed. There are other methods of attribute selection which are based on the transformation of input values providing information related to: how relevant is each variable as a whole?

It is possible to discard the ones that are irrelevant or those that are below a certain threshold of relevance. According to [9], filtering-based selection techniques are used to find the best subset of features of the original set. The filtering methods seem to be optimal for the selection of a subset of data. These do not depend on the classification algorithm and the computational cost is lower for large data sets. The wrapper-based choice features techniques (wrappers) also defined in [3], use the prediction capability of the classification algorithm to select the optimal subset of features. In this study the filtering-based selection techniques known as Info. Gain [4], Gain ratio [8] and Relief [5] have been used. In the references review, it was noticed that promising results can be got when applied to detect faults. During the experimental works carried out, Bayesian classifiers and networks were utilized to analyze the performance measurements obtained from the proposed models [4], [1]. The following is a detailed description of both the features selection techniques Info. Gain, Gain ratio, Relief and the classification methods Naïve Bayes, Bayesian networks which are based upon the suggested model.

3.1 Info.Gain

As presented in [2], Info. Gain is a filter-based features choice method. It is also known as information gain and is used to identify the importance of the features of a data set collection. The attribute with the largest gain is chosen as the division feature for the umpteenth node. This trait minimizes the required information to classify the couples in the resulting allocation. It reflects very small defects among these partitions.

3.2 Gain Ratio

As studied in [8], Gain ratio belongs to the category of filtering-based traits selection techniques which is applied to analyze the features of big size data sets. When there are many different values, the gain information relationship is used to consider these features. This approach is widely applied due to the good results that can be obtained. Additionally, these results can be employed during the classification phase. Its main distinctive feature is the modification of the information gain that reduces the error. Gain ratio considers the number and size of branches to choose from a characteristic.

3.3 Relief

This is an algorithm that determines significant features and allow to easily distinguish between instances of diverse classes [5]. Based on this approach, it defines the weight for each feature. However, the Relief genuine version limits its application field to two-class problems. Hence, for weight allocation purposes just the closest neighbor of a different class is utilized.

4 Classification and Training Techniques

At this stage, firstly, the classifier is trained. This process is done from the learning algorithm chosen and by using the normalized data set which is reduced to the most important features. Hence, an efficient learning is created. Once the training is performed, the classifier determines normal traffic and attacks through a subsequent classification of every connection within the data set. Then, the quality measures are computed to assess the classification technique performance. Bayesian classifiers were used. These are based on the Bayes theorem [3].

$$p(A|B) = \frac{p\,(A,B)}{p\,(B)} = \frac{p\,(A)\,p\,(B|A)}{p\,(B)} = \frac{p\,(A)\,p\,(B|A)}{\Sigma A' p\,(A')\,p\,(B|A')} \tag{2}$$

Being A and B two random events whose possibilities are denoted as $p(A)$ and $p(B)$ respectively and taking $p(B) > 0$. The A and B event possibilities previously known are supposed to be true. Likewise, the probability is subjected to event B to be true assuming that A is too $p(B|A)$. Finally, $p(A|B)$ is the possibility of A to be true considering that B is also true.

4.1 Naïve Bayes

As stated in [3], Naïve Bayes is a descriptive and predictive classification technique based upon the probability theory which comes from the Byes theorem analysis. This theory suggests both an infinite sample and an independent statistics among variables. In this case referring to the characteristics not to the class. Under these conditions, the probability distributions of each class can be calculated to establish the relationship between the traits and the class. If $x = (x_1, ..., x_n)$ is given, where x_i is the observed value for the umpteenth feature. Hence, the possibility for a class ym with k possible values $(y1, ..., yk)$ to occur, results from the Bayes rules as shown in equation 3.

$$P\,(y_m|x_1, x_2, \ldots, x_n) = \frac{p\,(y_m)\prod_{i=1}^{n} p\,(x_i|y_i)}{p\,(x_1, x_2, \ldots, x_n)} \tag{3}$$

In the above equation, $p(ym)$ is the class proportion of ym in the data set. Also, $p(x_i|y_i)$ is computed from the examples amount with a x_i value whose class is ym. Therefore, it can be inferred that to compute $p(x_i|y_i)$ makes the x_i values be discrete. So if there is some continuous feature, it should be discretized in advanced. The assorting of a new class ““x”” is done by calculating the conditioned possibilities of each class and choosing the best option. If $Y = (y1, y2, ..., yk)$ is the current class data sets, it will be sorted with the class that satisfies equation 4.

$$\forall i \neq j, \quad P\,(y_i|x_1, x_2, \ldots, x_n) > P\,(y_j|x_1, x_2, \ldots, x_n) \tag{4}$$

Although the bayesian classifier is a fast and simple method, it is required to go all over the training set to compute $P(y_m|x_1, x_2, ..., x_n)$. This calculation is unfeasible for a large number of examples. So, a simplification is needed. Therefore, the conditional independence hypothesis is considered for decomposing purposes of the probability.

4.2 Bayesians Networks

As stated in [1], a Bayesian network is a defined, directed and labeled acyclic graph, which describes the joint probability distribution that governs a set of random variables. Let $X = x1, x2, ..., xn$ be a set of random variables, a Bayesian Network for X is a pair $B =< G, T >$ in which:

- G is a directed acyclic graph in which every node represents a variable $x_1, x_2, ..., x_n$ and every arc symbolizes direct dependence relationships among the variables. The arcs direction shows that the variable pointed by the arc depends on the variable placed at its origin.
- T is a parameters set to quantify the network. It contains the probabilities $PB(xi|Xi)$ for each possible x_i value of each variable X_i and each possible value of n which denotes the parents set of X_i in G. Hence, a Bayesian network B defines a joint probability distribution over X as given in [1] and as indicated in equation 5.

$$P_B(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P\left(X_i | \prod x_i\right) \tag{5}$$

5 Methodology

The proposal in this work was initially to take the NSL-KDD (raw data) data set and to apply the pre-processing techniques: load balancing by data instances (through the implementation of Synthetic Minority Oversampling Technique - SMOTE) and normalization (applying standardization to zero average and unit variance). When the purified data is obtained a series of features selection techniques are applied to identify the attributes that affect the performance of the classifier. After filtering data, two Bayesian classification techniques are employed. The test process was performed through cross-validation using 10-folds. The results got from it were represented in the respective confusion matrices allowing the calculation of the quality metrics of each of the experimental scenarios. From this, the techniques (deselection and classification) which provide the best results were identified (Figure 1).

6 Simulations and Results

Two sets of experimental tests are involved in the development of this research. For the first set of tests, the Naïve Bayes classifier is used to change the features selection techniques (Info. Gain, Gain ratio and Relief). Once the corresponding feature selection technique is applied, the priority order of the attributes can be identified. Based on this, a series of experimentation scenarios are simulated in which the number of attributes for each of the selection techniques implemented are varied. See Table 2 and figures 2(a), 2(c) and 2(e). For the second set of tests the Bayesian networks classifier is applied, and the features selection techniques are also varied. After identifying the priority order of the attributes, the experimental scenarios are carried out. In these simulations the traits number are modified for each of the choice techniques implemented. (See table 3 and figures 2(b), 2(d) and 2(f). For the set of tests carried out with the Naïve Bayes classifier, the best results have been obtained by using the selection technique of Relief features with 20 attributes. An accuracy of 91.27% was reached as shown in table 2. For the tests carried out with Bayesian Networks, the best results are obtained with Gain ratio using only 13 attributes with a success rate of 97.56% (See Table 3). The most significant simulation scenarios are provided in Table 3. The compilation described does not intend to indicate that Bayesian Network + Gain ratio is the best solution. The goal is to give a performance perspective of the proposed procedure compared to the results provided by previous studies.

The methods shown in table 5 have been classified as: (1) Methods that do not use features selection, (2) filtering-based methods and (3) wrapping-based methods. All the experimental work is performed by using a MacBook Air mid 2015 with an Intel processor, 1.6 Ghz and an 8 gb RAM DDR3 at 1600Mhz. Each experiment is completed 10 times. Thanks to this, metrics values are obtained, and it allows to evaluate the quality of the processes. See tables 2 and 3 in which each quality metric by technique of selection of features and by classification method is shown with their respective standard deviation. In the classifying process of both sets of tests, the crossed to 10 folds validation technique is used. It is applied to the NSL KDD data set of training. Simulations that allow to evaluate the network traffic with a behavior like real computer attacks are generated. When each experimental scenario is solved, the evaluation of the proposed functional models is performed. Hence, the metrics of accuracy, sensitivity and specificity can be computed.

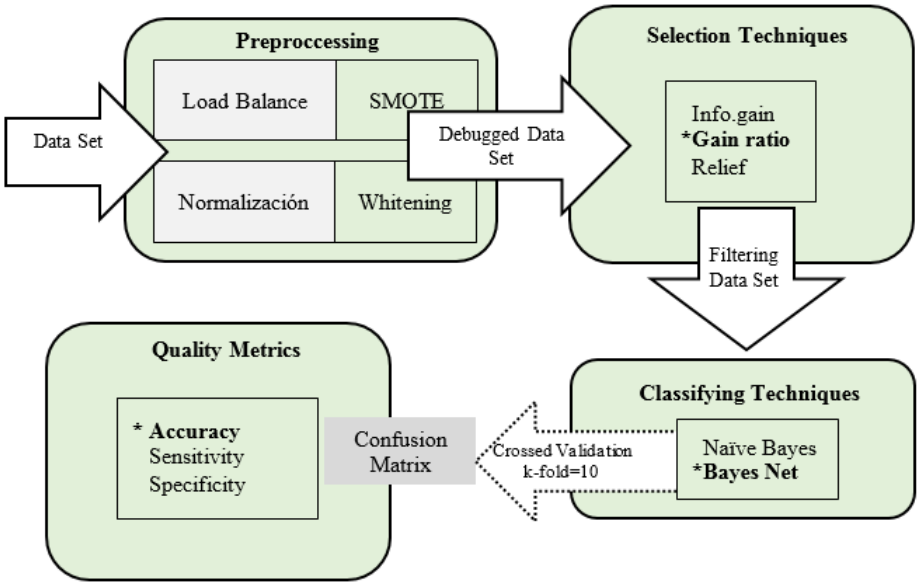
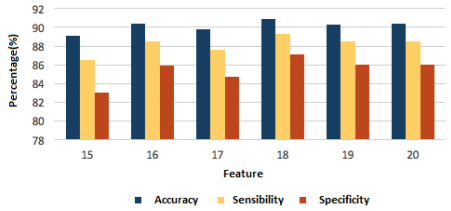


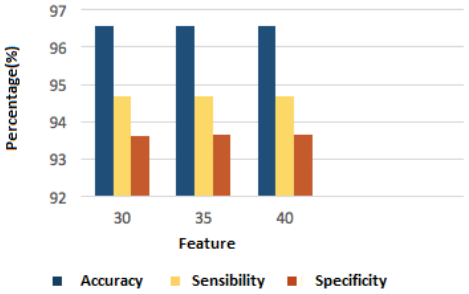
Fig. 1: Suggested Methodology

7 Related Works

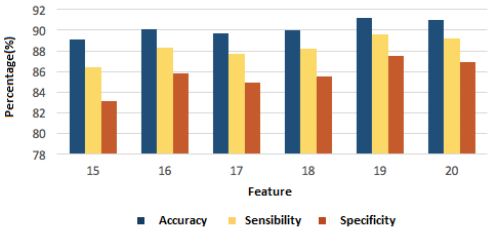
In [16] an analysis of the features selection techniques for a data set of network traffic like the one proposed here was done. The Naïve Bayes, J48 and PART classifiers were utilized. The performance of each of these classifiers was assessed with the entire data set NSL-KDD and with subsets of data identified from the application of different features selection techniques. The best results were obtained from the PART classifier (97.57% of accuracy). The techniques Chi Square (30 features) and Consistency (14 features) were individually applied. The results acquired were very similar to the ones got with Bayesian network + Gain ratio (97.56%) and just 13 attributes were used. See Table 3. Moreover, in this study, tests with Naïve Bayes and Gain ratio feature selection techniques are also performed (89.03%) and Info. Gain (93.49%). Both tests with 30 features. In the scenarios set up for this research with Naïve Bayes experimentation + Gain ratio + 19 features, a correct rate is obtained 91.22%. Naïve Bayes + Info. Gain + 16 features with a success rate of 90.41%. In [20], a combination approach to feature selection techniques is proposed for intrusion detection systems. In this work the number of attributes is reduced by using different classification techniques based on feature selection and evaluation is done through ten classification algorithms that generate the most representative results. The best results are achieved with the IBK classifier and the combination of selection techniques such as Symmetric and Gain ratio with 15 features reaching a success rate of



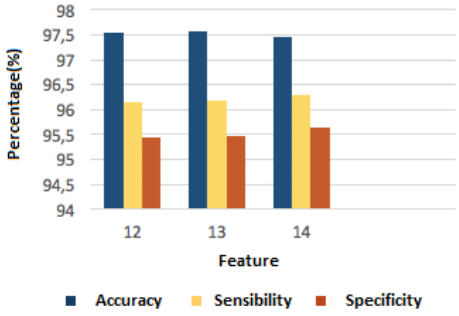
(a) Naïve Bayes - Info.Gain



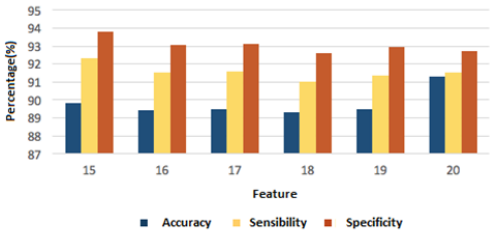
(b) Bayesian Networks - Info.Gain



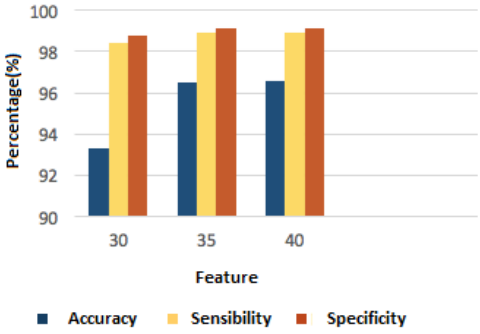
(c) Naïve Bayes Gain Ratio



(d) Naïve Bayes - Gain Ratio



(e) Naïve Bayes - Relief Ratio.



(f) Bayesian Networks - Relief

Fig. 2: Tests Set Results for Naïve Bayes and Bayesian Networks

98.5%. In [3], a method based on wrapper applied with a multi-objective approach using the GHSOM classifier is studied. It is employed with a probabilistic adaptation for the re-labeling process allowing to differentiate between normal and anomalous traffic as well as different typologies of anomalies. This proposal provided a rate of 99.12 ± 0.61 in which 25 features are analyzed.

Table 2: Naïve Bayes Classifier Tests

Selection Technique	Features	Percentage (%)	Sensitivity (%)	Specificity (%)
Info.Gain	15	89.14 ± 0.35	86.47 ± 0.35	83.070 ± 0.43
	16	90.41 ± 0.35	88.46 ± 0.35	85.91 ± 0.42
	17	89.80 ± 0.37	87.61 ± 0.37	84.74 ± 0.43
	18	90.86 ± 0.41	89.34 ± 0.43	87.13 ± 0.32
	19	90.30 ± 0.35	88.48 ± 0.37	85.97 ± 0.37
	20	90.35 ± 0.43	88.54 ± 0.43	86.05 ± 0.47
Gain ratio	15	89.07 ± 0.36	86.460 ± 0.43	83.090 ± 0.82
	16	90.09 ± 0.52	88.35 ± 0.65	85.84 ± 0.4
	17	89.67 ± 0.47	87.72 ± 0.48	84.96 ± 0.77
	18	90.03 ± 0.39	88.17 ± 0.47	85.57 ± 0.78
	19	91.22 ± 0.41	89.63 ± 0.73	87.48 ± 0.42
	20	90.95 ± 0.43	89.21 ± 0.81	86.92 ± 0.49
Relief	15	89.8 ± 1.1	92.32 ± 0.76	93.81 ± 0.78
	16	89.45 ± 0.97	91.50 ± 0.93	93.08 ± 0.87
	17	89.49 ± 0.81	91.56 ± 0.84	93.14 ± 0.76
	18	89.30 ± 0.93	90.99 ± 0.74	92.61 ± 0.84
	19	89.47 ± 0.75	91.36 ± 0.59	92.94 ± 0.67
	20	91.27 ± 0.82	91.50 ± 0.53	92.74 ± 0.71

Table 3: Bayesian Networks Tests

Selection Technique	Features	Percentage (%)	Sensitivity (%)	Specificity (%)
Info.Gain	–	–	–	–
	30	96.55 ± 1.02	94.670 ± 0.93	93.61 ± 0.73
	35	96.56 ± 1.21	94.690 ± 0.51	93.64 ± 0.49
	40	96.56 ± 1.02	94.690 ± 0.47	93.64 ± 0.51
	–	–	–	–
Gain ratio	–	–	–	–
	12	97.55 ± 0.41	96.14 ± 0.36	95.44 ± 0.47
	13	97.56 ± 0.53	96.17 ± 0.51	95.47 ± 0.49
	14	97.46 ± 0.56	96.29 ± 0.70	95.63 ± 0.53
	–	–	–	–
Relief	–	–	–	–
	30	93.30 ± 0.92	98.40 ± 0.98	98.77 ± 0.83
	35	96.54 ± 1.03	98.92 ± 0.75	99.11 ± 0.97
	40	96.56 ± 1.28	98.93 ± 0.83	99.12 ± 1.03
	–	–	–	–

8 Conclusions

The success rate is the most appropriate metric to evaluate the performance of a classifier (regarding the level of traffic detection in computer networks). It can be verified that better values are obtained with the Bayesian network classifier with the Gain ratio feature selection. Some of the features that best contribute to the classification process are: logged_in, srv_serror_rate, flag, serror_rate, dst_host_srv_serror_rate, diff_srv_rate, dst_host_serror_rate, dst_host_srv_diff_host_rate and wrong_fragment. The quality metrics obtained is: a success rate of 97.56%, 96.17% of sensitivity and the specificity of 95.47%. Using the thirteen most relevant characteristics of the 41 possible attributes of the NSL-KDD dataset helps to create a lighter IDS.

An important improvement in the detection rate of attacks and normal traffic in computer networks has been identified. A lower proportion of features and less computational resources are ap- plied. It may be useful for a later solution on equipment with lower performance and if necessary, for a real time analysis. an exhaustive comparison would be required which is currently not possible because the only available performance results refer to the success rate. Most of the the similar research works did not present statistical significance test from which the standard deviation of the success rates could be extracted. Further, there are not specific implementations to be able to execute and compare the results obtained in a more detailed way.

9 Future Work

As future research work, an exhaustive review combining features selection techniques with different classifying techniques is proposed. So, this will allow to determine the optimal number of characteristics to acquire the best results and the appropriate classifier. In addition, a choice technique based upon wrapper will be developed in which an optimal classification technique is integrated and identified from the experimental processes suggested.

References

[1] Hota H.S. Shrivas A.K. "Data Mining Approach for Developing Various Models Based on Types of Attack and Feature Selection as Intrusion Detection Systems (IDS)". In: Conference paper on Intelligent Computing, Networking and Informatics. Springer, 2013, pp. 845–851.

[2] De-La-Hoz-Franco Emiro Ortiz Andrés Ortega Julio De-La-Hoz-Correa Eduardo Prieto Alberto. "Network anomaly detection with bayesian self-organizing maps". In: Lecture Notes in Computer Science - IWAN Conference. Vol. 7902. Tenerife - Spain: Springer, 2013, pp. 530–537.

[3] De-La-Hoz-Franco Emiro De-La-Hoz-Correa Eduardo Ortiz Andrés Ortega Julio Martínez-Alvarez Antonio. "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps". In: *Knowledge-Based Systems* 71 (2014), pp. 322–338.

[4] De-La-Hoz-Franco Emiro De-La-Hoz-Correa Eduardo Ortiz Garcia Andrés Ortega Lopera Julio Martinez-Alvarez Antonio. "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps". In: *Knowledge-Based Systems* 71 (Aug. 2014), pp. 322–338.

[5] De-La-Hoz-Correa Eduardo De-La-Hoz-Franco Emiro Ortiz Andrés Ortega Julio Prieto Beatriz. "PCA filtering and probabilistic SOM for network intrusion detection". In: *Neurocomputing* 164 (N/A 2015), pp. 71–81.

[6] Bazara B. Anthony C.H. "Chapter 10. Intrusion Detection Systems". In: *Handbook of Information and Communication Security*. New York: Springer-Verlag, 2010, pp. 193–215.

[7] Sun Yijun Wu Dapeng. "A RELIEF Based Feature Extraction Algorithm". In: Proceedings of the 2008 SIAM International Conference on Data Mining. Atlanta-Georgia: Society for Industrial and Applied Mathematics, 2008, pp. 188–195.

[8] De-La-Hoz-Franco Emiro Ortiz Garcia Andrés Lopera Ortega Julio De-La-Hoz-Correa Eduardo Mendoza-Palechor Fabio. "Implementation of an Intrusion Detection System based on Self Organizing Map". In: *Journal of Theoretical and Applied Information Technology - JATIT* (Jan. 2015), pp. 324–334.

[9] De la Hoz Eduardo Ortiz Andrés De la Hoz Emiro Ortega Julio. "Network Anomaly Classification by Support Vector Classifiers Ensemble and Non-linear Projection Techniques". In: Hybrid Artificial Intelligent Systems. 8th International Conference, HAIS. Ed. by Heidelberg. Springer Berlin. Salamanca: Lecture Notes in Computer Science, 2013, pp. 103–111.

[10] Bayes T. Price R. Canton J. *An essay towards solving a problem in the doctrine of chances*. Royal Society of London, 1763.

[11] Karegowda A.G. Manjunath A.S. Jayaram M.A. "Comparative study of attribute selection using Gain ratio and Correlation based feature selection". In: *International Journal of Information Technology and Knowledge Management* 2 (2 Dec. 2010), pp. 271–277.

[12] Li Kang Deolalikar Vinay Pradhan Neeraj. "Big Data Gathering and Mining Pipelines for CRM using Open-source". In: 2015 IEEE International Conference on Big Data. Ed. by IEEEXplore. IEEE, 2015, pp. 2936–2938.

[13] None. *Instituto Nacional de Ciberseguridad de España*. June 2017. URL: https://www.certsi.es/respuesta-incidentes.

[14] Karna Nyoman Supriana Iping Maulidevi Nur. "Social CRM using web mining for Indonesian academic institution". In: Information Technology Systems and Innovation (ICITSI), 2015 International Conference on. Ed. by IEEEXplore. Vol. 15. Bandung, Indonesia: IEEE, 2015, pp. 1–6.

[15] Atkinson Malcolm Baxter Rob Brezany Peter Corcho Oscar. "Analytical Platform for Customer Relationship Management". In: Atkinson Malcolm Baxter Rob Galea Michelle Parsons Mark Brezany Peter Corcho Oscar Hemert Jano Snelling David. *The DATA Bonanza: Improving Knowledge Discovery in Science, Engineering, and Business*. Ed. by Sons John Wiley &. Vol. 1. Scotland: Wiley-IEEE Computer Society Press, 2013, pp. 287–300.

[16] Singh Raman Kumar Harish Singla R.K. "Analysis of Feature Selection Techniques for Network Traffic Dataset". In: *International Conference on Machine Intelligence aRnedseRaercsheaarnchd Advancement* (15 2013). Ed. by Society IEEE Computer, pp. 42–46.

[17] Ummugulthum S. Baulkani S. “Customer Relationship Management Classification Using Data Mining Techniques”. In: International Conference on Science, Engineering and Management Research. Ed. by IEEEXplore. IEEE, 2014, pp. 27–29.

[18] Rotovei Doru Negru Viorel. “A methodology for improving complex sales success in CRM Systems”. In: INnovations in Intelligent SysTems and Applications (INISTA), 2017 IEEE International Conference on. Ed. by IEEEXplore. Vol. 17. IEEE, 2017, pp. 1–6.

[19] Ippoliti Dennis Zhou Xiaobo. “A-GHSOM: An adaptive growing hierarchical self organizing map for network anomaly detection”. In: *Journal of Parallel and Distributed Computing* 72 (12 Dec. 2012), pp. 1576–1590.

[20] Garg Tanya Kumar Yogesh. “Combinational Feature Selection Approach for Network Intrusion Detection System”. In: *nternational Conference on Parallel, Distributed and Grid Computing* (2014), pp. 82–87.