Feature Selection with IG-R for Improving Performance of Intrusion Detection System

Yakub Kayode Saheed[1], Fatimah Enehezei Hamza-Usman[2]

[1]Department of Computer Science, Cyber Security Unit, Al-Hikmah University, Ilorin, Nigeria
[2]Department of Computer Science, University of Ilorin, Ilorin, Nigeria

Abstract**:** As the popularity of the internet computer continued to grow and become an indispensable in human life, the security of computer network has become an important issue in computer security field. The Intrusion Detection System (IDS) is a system used in computer security for network security. The feature selection stage of IDS is considered to be the most critical stage in IDS. This stage is very costly both in efforts and time. However, many machine learning approaches have been presented to improve this stage in order to improve the performance of an IDS. However, these approaches did not give desirable results with respect to the detection accuracy in the IDS. A novel technique is proposed in this paper combining the Information Gain and Ranker (IG+R) method as the feature selection strategy with Naïve Bayes (NB), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) as the classifiers. The performance of these IG+R-NB, IG+R-SVM, and IG+R-KNN was evaluated on NSLKDD dataset. The experimental results of our proposed method gave high accuracy and low false alarm rate. The results obtained was compared and benchmarked with existing works. The results of this paper outperformed the existing approaches in terms of the detection accuracy.

Keywords**:** Network Security, Intrusion Detection System, Information Gain, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor

1. Introduction

Internet has brought huge potential for business and has a profound impact on people's live [1]. However, it poses lots of security concerns in terms of the risks and threats.

Intrusion detection is a network security method for preventing, avoiding, and stopping an illicit access to a computer network[2]. Intrusion Detection Systems (IDS) perform a significance function in achieving a protected and secured network. The best way to measure the effectiveness of an IDS centered on how successful it is in maximizing its detection accuracy while minimizing the false alarm rate. Anomaly-based IDS have been an active area of research in IDS field simply because of their success in recognizing and unaware attacks. The best way to measures the effectiveness of an IDS centered on how successful it is in maximizing its detection accuracy while minimizing the false alarm rate. Anomaly-based IDS have been an active area of research in IDS simply because of their success in recognizing an unaware attacks[2].

In order to avert and avoid attacks on network, a Network Intrusion Detection System (NIDS) might subsist with machine learning classifiers to improve the accuracy and detection speed. The application of machine learning has additional merit that expert knowledge is not required as much as the white list or black list model[3].

The IDS system is categorized into two based on the detection. The anomaly detection and the signature detection, the formal detection compares all behavior against the normal defined activity while the latter spot the traffic pattern as malicious and this requires an updated database for storing all the new attack signatures[4].

In the last decade, several efforts have been made by researchers in computer security field by using machine learning classifier for improving the IDS system. The researchers employed algorithms such as C4.5 [5], k-means clustering and Decision tree [6], average one dependence estimator[7], genetic algorithm[8], ID3 and random forest[9]. The detection accuracy and high false alarm rate are still major issues to address in IDS systems. Hence, this paper attempt to improve the detection accuracy and reduce the false alarm rate.

This paper is divided into three sections. Section 1 is the introduction. Section 2 presents the related work. Section 3 describes the methodology and section 4 shows the results and discussion. Section 5 concludes the paper.

2. Related work

Several approaches and methods have been reported in the literature for intrusion detection system. The approaches and techniques all used machine learning techniques.

The authors [10]in proposed an approach for IDS through feature selection analysis and hybrid efficient model. The experimental result in this work was 99.81% accuracy and 98.56% for the binary class and multiclass NSL-KDD datasets respectively. However, there are problems with low false negative rates and high false alarm rate.

[11] presented a hybrid intelligent technique using grouping of classifiers. They use 2-class classification approach with 10-fold cross validation technique to generate the final classification results with respect to intrusion or normal network. The experimental analysis was done on NSL-KDD dataset and results showed that the approach is efficient with high detection rate and low false alarm rate.

The study [12] aims to use data mining method classification tree and support vector machines for intrusion detection. The experiment was performed on KDD CUP 99 data and the results obtained showed that the C4.5 algorithm outperformed the SVM in the network intrusion detection and false alarm rate.

The [13] work is based on hybrid approach of GA and SVM for network intrusion detection systems. The proposed approach has the capacity to decrease the features of the dataset from 41 to 10. The features selected were distributed into three priorities based on GA where the highest significant placed in the first priority and the lowest significant in the third priority. The features were distributed by placing four features in the first priority, another four in the second priority and two features in the third priority. The experimental results of the hybrid technique gave 0.973 positive detection with false alarm rate of 0.017.

Another hybrid method was introduced by [14], they combine multiple classifiers for classifying normal and anomalous activities in the computer network. The C5.0 Decision tree classifier was used to construct the misuse detection model and one-class SVM is used to implement the anomaly detection model. The collection of multiple classifiers helps to improve the performance. The experimental analysis was carried out on NSL-KDD dataset and the results findings showed that the overall performance of the method is improved in terms of low false alarm rate and detection rate when compared to the existing method.

3. Methodology

This section provides information on our proposed approach. The algorithms used and the architecture of our proposed system is presented in this section. The application of machine learning in IDS can be categorize into three[15]: 1) feature selection, 2) pre-processing and 3) model training. The first category is considered to be the most critical because raw network traffic data are transformed in this stage, a stage that is very costly both in time and effort. The 2) category focused on the study of methods for feature selection in the dataset. And 3) deals with the performance of different classifiers for IDS in computer networks[16], [17], [18].

In this study, we focused and placed more emphasis on category 1 that is considered to be the most critical stage. We optimized the features in the dataset by eliminating the irrelevant and redundant features vis-a-viz improving the performance of our classifiers.

3.1 Feature Selection

Feature selection can be describe as eliminating the redundant and irrelevant attributes [19] from a dataset in order to optimize learning performance in terms of detection accuracy, false alarm and time to build the model. In respect of the selection approaches, feature selection can be of three types: filter, wrapper and embedded methods. The filter feature selection based on IG was employed in this paper. The block diagram of this study is shown in figure 1.

3.2 Information Gain

Information Gain (IG) can be define as an entropy-based feature evaluation technique that is extensively used in machine learning field and measures how much information a feature gives facts in respect of the target class [19].IG measures how features are mixed up [20], [21] and also depends on information entropy which the attributes provide to the model [22].In classification system, each individual feature in a feature vector is in connection with IG [23].IG is used to find the quantity of information gotten for group prediction by ascertaining if a feature is absence or presence [24].The IG is obtained by counting the amount of each feature occurrences in each category[25]. In IDS, IG is used to measure and find the relevance of feature J in class K. The more the value of mutual information between classes K and feature J, the more the relevance between classes K and feature J[20].

$$I(K,J) = H(K) - H(K|J) \quad (1)$$

Where H (K) = - $\sum_{kEK} p(K)$ log p(K), the entropy of the class, and H(K|J) is the conditional entropy of class given, feature H (K|J) = - $\sum_{kEK} p(K|J)$ log p(K|J).

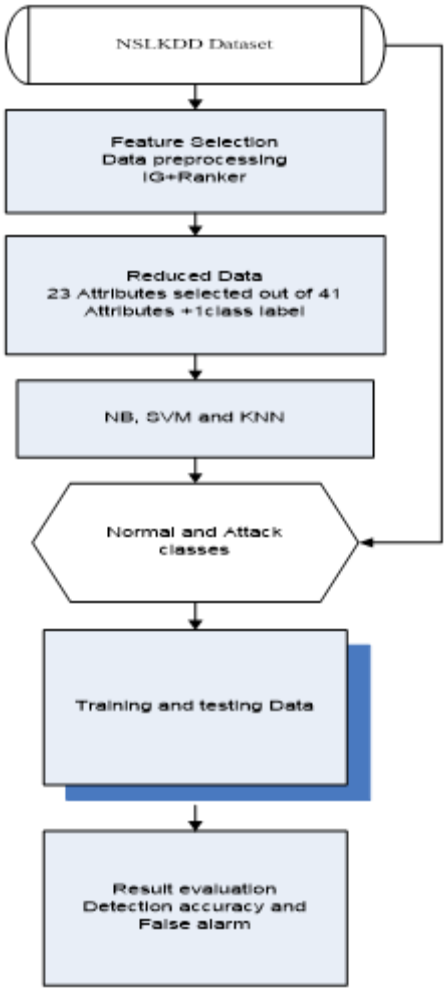

Figure 1. Block diagram of proposed IG+R-NB, IG+R-SVM and IG+R-KNN

The IDS dataset has balanced class, the probability of class K for both Normal and attack classes is equal to 0.5. Accordingly, the entropy of classes H (K) is equal to 1. Consequently, the IG can be expressed as

$$I(K, J) = 1 - H(K|J). \quad (2)$$

3.2.1 Proposed IG+R-NB, IG+R-SVM, and IG+R-KNN

As feature selection in IDS is a crucial part, we employed IG, it uses an Attribute Evaluator that evaluates the attributes and a ranker to rank all the features in the dataset. The number of features select from feature vector was defined to be 18 out of 41 features with 1 class label for all the attack classes in the dataset. We removed the features one at a time with lower rank from the bottom of the ranking and observe the weight put by the ranker algorithm. The IG+R filter approach was carried out to eliminate both the redundant and irrelevant features in the NSLKDD dataset in order to improve the performance of the model. The 18 features filtered out with IG is performed on all the attack classes and Normal class as show in Table 1.

Table 1. Filtered features using IG and Ranker search method

Class Category	Filtered Features
U2R	16,35,20, 18, 22, 23, 24, 25, 34, 17, 18, 26, 27, 32, 31, 36, 5, 3
R2L	13, 16, 15, 14, 25, 17, 22, 18, 24, 23, 12, 26, 10, 5, 37, 38, 6, 4
Probe	17,8, 14, 6, 5,15, 18, 4, 12, 11, 10, 16, 13,19,7, 23, 36, 1
DoS	17,6,18,4,15, 8,12,11,14,16, 13,19, 10, 7, 5, 25, 24, 37
Normal	11,17,6, 8,15,18, 4, 12,14, 19,16,13, 10,5,7, 1, 25, 37

3.2.2 Naïve Bayes

NB is a powerful and easy to build classifier, with no complex iterative parameter approximation which made it a good fit for huge datasets[26]. It is used to predict the probability of a class to belong to either normal or attack classes. It performs easily in both training and classification stages [27]. NB assumption is that all features in the feature vector are equally independent and important[28].

NB works as follows.

1. Let T be a training set of samples, each with their class labels. There are k classes, H_1, H_2, H_3,…..H_k. Each sample is represented by an n-dimensional vector X = {X_1, X_2, X_3,…..X_n },depicting n measured values of the n features, F_1, F_2, F_3,…..F_n ,respectively.
2. Given a sample X, the classifier will predict that X belongs to the class having the highest a posteriori probability, conditioned on X.

The NB theorem is given as

$$P(H/X) = P(H/X)P(H)/P(X) \quad (3)$$

Where X-Tuples, H-Hypothesis, P (H/X) represents posterior probability of H conditioned on X.

NB classifiers simplify the computations and exhibit high accuracy and speed when applied to large dataset.

3.2.3. Pseudocode of Naïve Bayes Algorithm

Input:

Training dataset T,
F= (f1, f2, f3,.., fn) // value of the predictor variablein testing dataset.

Output:

A class of testing dataset.

Step:

1. Read the training dataset T;
2. Calculate the mean and standard deviation of thepredictor variables in each class;
3. Repeat

Calculate the probability of *fi* using the gaussdensity equation in each class;

Until the probability of all predictor variables (f1, f2,f3,.., fn) has been calculated.

4. Calculate the likelihood for each class;
5. Get the greatest likelihood.

3.2.4. Support Vector Machine

SVM is a general, popular and useful classifier[29]. SVM gives good generalization power, robust against local minima and represented by small parameters [30].The principle of SVM is to construct a hyperplane to ensure that the distance between the two types of structure is maximized[31], The SVM hyperplane can be given as.

$$w.x + b = 0, \quad (4)$$

where w means the weight vector, x means the input dataset, and b means a bias constant in the hyperplane.

3.2.5. K-Nearest Neighbor

K-NN method to classification is a completely non-parametric [32], [33]and an instance based learning method for classifying objects based on the closest training examples in the feature space. This a type of lazy learning algorithm where all the computations are delayed to classification stage and the function is approximated locally. The KNN is one of the simplest classifiers in machine learning. The k-NN algorithm uses all labeled training instances as a model of the target function. An advantage of the K-NN Algorithm as a classifier for an IDS is that it is analytically tractable. The Euclidean distance is given as.

$$d(X,Z) = \sqrt{\sum_{i=1}^{n}(Zi - Xi)2} \quad (5)$$

3.2.6.K-Nearest Neighbour Algorithm Pseudocode

Let (X_i, C_i) where i = 1, 2……., n be data points. X_i denotes feature values and C_i denotes labels for X_i for each i.
Assuming the number of classes as 'c'
$C_i \in \{1, 2, 3, \ldots\ldots, c\}$ for all values of i

Let x be a point for which label is not known, and we would like to find the label class using k-nearest neighbor algorithms.

Procedure:

1. Calculate "d(x, x_i)" i =1, 2, ….., **n**; where **d** denotes the Euclidean distance between the points.
2. Arrange the calculated **n** Euclidean distances in non-decreasing order.
3. Let **k** be a +ve integer, take the first **k** distances from this sorted list.
4. Find those **k**-points corresponding to these **k**-distances.
5. Let $\mathbf{k}_i$ denotes the number of points belonging to the i^{th} class among **k** points i.e. $k \geq 0$
6. If $k_i > k_j \; \forall i \neq j$ then put x in class i.

4. Results and Discussion

The goal of our experiment is to show how the three classification algorithms can efficiently and effectively able to detect intrusions. We used three classification algorithms Naïve Baye, SVM and KNN. The benchmark dataset used in this research is the NSL-KDD dataset. Next, we will discuss about the data used to train and test the classifiers.

4.1. Description of the NSLKDD Data Set

The NSLKDD dataset is an improvement over the KDD99datast. This dataset consists of four attack class categories known as U2R, R2L, probe and DoS which is made up of 41 features and 1 normal class label. In this dataset, three main problems were solved. The first problem is that the observations that are repetitive in the testing and training sets were removed to eliminate biasing classification methods towards the most repeated observations. Secondly, the testing and training set were created by picking out observations from different parts of the original KDD99 dataset. And thirdly, the observations that was notice to be

imbalanced in each attack class either in the testing set or training set were addressed to reduce the FAR[33].
The distribution and spreading of attack classes and normal class record in the NSLKDD dataset for both training and testing set is shown in Table 2.

Table 2. NSLKDD Data set

Class Category	Training set	Testing set
U2R	52	67
R2L	995	2,887
Probe	11,656	2,422
DoS	45,927	7,458
Normal	67,343	9,710
Total records	125,973	22,544

The experiments were carried out on a 64-bit Windows 10 Professional operating system, x64-based processor with 8.00 GB of RAM and Intel (R) Core (TM)i5-8250U CPU @1.60 GHz 1.80GHz.

4.2 10-fold Cross Validation

In order to assess the effectiveness of the algorithms, each one of them was trained on the NSLKDD data set using a ten-fold validation test mode. To test and evaluate the algorithms, we use 10-fold cross validation[34][36]. The data set is separated into 10 subsets in this process [35]. For each of the time, one out of the 10 subsets served as the test set whereas the remaining k-1 subsets constitute the training set. The statistical performance is calculated across the 10 trials. This offers a good suggestion of how fit the classifier will do on unseen data.

4.3 Performance Measurement Terms

(1). Correctly Classified Instance: The correctly and incorrectly classified instances demonstrate the percentage of test cases that were correctly or properly and erroneously or incorrectly classified. The percentage of correctly classified instances is often called accuracy. Accuracy is the most important metric in intrusion detection system [37]. We based the performance of our proposed model on accuracy and false alarm.
(i). True positive (TP): It connotes the correctly rejected, and it indicates the number of anomaly records that is recognized as anomaly.
(ii). False Positive (FP)or false alarm: is the number of incorrectly rejected, and it connotes the number of normal records that are recognized as anomaly.
(iii). True Negative (TN): is equal to those records correctly admitted, and it connotes the number of normal records that are recognized as normal.
(iv). False Negative (FN): is equal to the records that are incorrectly admitted, and it connotes number of anomaly records that are recognized as normal. The confusion matrix is shown in Table 3.

Table 3. Confusion Matrix

Predicted Class / Actual Class	anomaly	normal
anomaly	TP	FN
normal	FP	TN

The performance of an IDS is to have high accuracy, high detection rate and with lower false alarm rate [33].We employed these metrics as performance measure in this paper with respect to the confusion matrix [42].

4.4 Experimental results of NB classifier

Although NB are capable of handling a 5-class classification problem. We built five (5) different classifiers. The dataset is divided into two (2) classes of "Normal" and "Attack" patterns where Attack means the group of four classes (Probe, DoS, U2R, and R2L) of attacks. The aim is to detached normal and attack patterns. The process is performed and repeated for all the five (5) classes. Firstly, a NB classifier was built utilizing the training data and the testing data was tested with the built classifier to classify and categorize the data into normal class or attack classes. The performance of NB is shown in table 4 and table 5.

4.5 Experimental results of SVM

SVM is capable of binary class classification problems, we used SVM five times for detecting the attacks type. The SVM classifier learns from the training set of data and is also used on the test data set to classify and categorize the data into normal class or attack classes pattern. This process is performed and repeated for all the classes. The results are shown in **Table 6 and 7.**

Table 4. Performance evaluation of NB

Parameters	NORMAL	U2R	R2L	PROBE	DoS
Correctly Classified Instances (%)	90.6652	94.1753	95.1896	89.5285	96.1072
Incorrectly Classified Instances (%)	9.3348	5.8247	4.8104	10.4715	3.8928
Kappa Statistics	0.8117	0.016	0.3555	0.6655	0.9193
Mean Absolute Error	0.0936	0.0594	0.052	0.1049	0.0395
Root Mean Squared Error	0.2981	0.2103	0.2117	0.3186	0.1927
Relative Absolute Error	18.8065	3461.8451	172.2511	42. 1876	8.187
Root Relative Squared Error	59.7531	735.8959	172.4906	90.372	39.2262

Table 5. Performance measurement of NB

Parameters	Normal	U2R	R2L	Probe	DoS
TP Rate	0.907	0.942	0.952	0.895	0.961
FP Rate	0.098	0.364	0.071	0.057	0.042
Precision	0.908	1	0.987	0.932	0.961
Recall	0.907	0.942	0.952	0.895	0.961
F-Measure	0.906	0.97	0.966	0.905	0.961
ROC Area	0.965	0.87	0.954	0.963	0.973

Table 6. Performance evaluation of SVM

PARAMETERS	NORMAL	U2R	R2L	PROBE	DOS
Correctly Classified Instances (%)	97.4197	99.9183	99.2898	99.1613	98.7303
Incorrectly Classified Instances (%)	2.5803	0.0817	0.7102	0.8387	1.2697
Kappa Statistics	0.948	0	0.6964	0.9658	0.9736
Mean Absolute Error	0.0258	0.0008	0.0071	0.0084	0.0127
Root Mean SquaredError	0.1606	0.0286	0.0843	0.0916	0.1127
Relative Absolute Error	5.1844	47.6408	23.5067	3.3736	2.6302
Root Relative Squared Error	32.2005	100.0398	68.6532	25.9773	22.9354

Table 7. Performance measurement of SVM

PARAMETERS	NORMAL	U2R	R2L	PROBE	DOS
TP RATE	0.974	0.999	0.993	0.992	0.987
FP RATE	0.029	0.999	0.452	0.038	0.018
PRECISION	0.975	0.998	0.993	0.992	0.988
RECALL	0.974	0.999	0.993	0.992	0.987
F-MEASURE	0.974	0.999	0.992	0.992	0.987
ROC AREA	0.973	0.5	0.77	0.977	0.985

4.6. Experimental results of KNN classifier

KNN can be used to handle binary classification problems, five KNNs was constructed for detecting the attack type. The KNN classifier learns from the training set of data and is also used on the test data set to classify and categorize the data into normal class or attack classes pattern.

This process is performed and repeated for all the classes. The results are shown in **Table 8 and table 9.**

Table 8. Performance evaluation of KNN

Parameters	Normal	U2R	R2L	Probe	DOS
Correctly Classified Instances (%)	98.9294	99.896	99.7071	99.7204	99.6517
Incorrectly Classified Instances (%)	1.0706	0.104	0.2929	0.2796	0.3483
Kappa Statistics	0.9785	0.3631	0.9024	0.9887	0.9928
Mean Absolute Error	0.0108	0.0011	0.003	0.0029	0.0035
Root Mean Squared Error	0.1035	0.0322	0.0541	0.0529	0.059
Relative Absolute Error	2.1641	64.0584	9.9608	1.1523	0.7315
Root Relative Squared Error	20.7403	112.851	44.0829	14.9969	12.0117

The results of our experiments showed that the SVM gives accuracy of 99.9% for U2R attack class, KNN gives 99.8% for U2R attack class and NB gives 94% for U2R attack class which gives less accuracy for detecting any class of attack. For Normal class, the KNN gives 98.9%, SVM gives 97% and NB gives 90%. These experimental results showed that the best algorithm for detecting the attack is SVM followed by KNN that performed better for detecting the Normal class as depicted in figure 2.

Table 9. Performance measurement of KNN

PARAMETERS	NORMAL	U2R	R2L	PROBE	DOS
TP RATE	0.989	0.999	0.997	0.997	0.997
FP RATE	0.011	0.6336	0.099	0.011	0.004
PRECISION	0.989	0.999	0.997	0.997	0.997
RECALL	0.989	0.999	0.997	0.997	0.997
F-MEASURE	0.989	0.999	0.997	0.997	0.997
ROC AREA	0.99	0.796	0.952	0.993	0.997

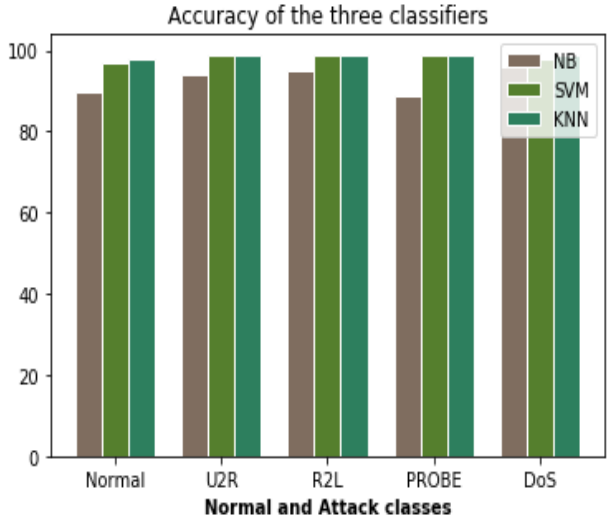

Figure 2. Accuracy of the IG+R-NB, IG+R-SVM and IG+R-KNN

In terms of false alarm rate, KNN gives 0.4% for DOS attack class, SVM gives 1.8% for DOS attack class and NB gives 4.2% for DOS attack class. However, for the Normal class, KNN gives 1.1%, SVM gives 2.9% and NB gives 9.8% for the Normal class as shown in figure 3.

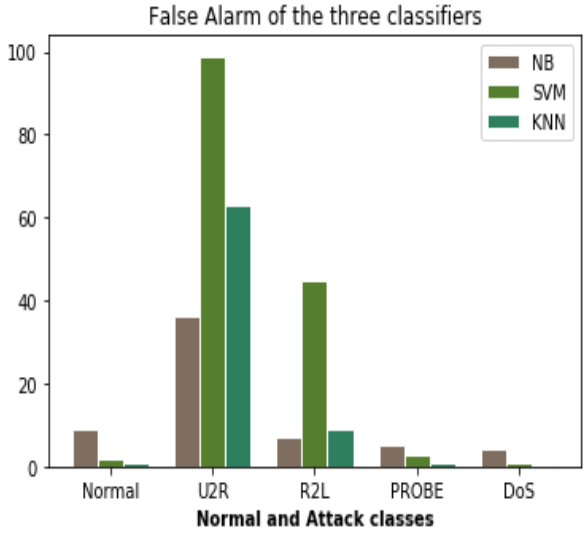

Figure 3. False alarm of the IG+R-NB, IG+R-SVM and IG+R-KNN.

The detection accuracy obtained from the proposed study was benchmark against existing works. The experimental results from our findings outperformed the existing work with respect to the detection accuracy as shown in Table 10.

Table 10. Comparison with existing works

Authors	Year	Detection accuracy (%)
[38]	2019	85.5
[39]	2018	91.3
[40]	2018	95.7
[41]	2017	79.7
Proposed IG+R-NB	**2020**	**99.8**
Proposed IG+R-SVM	**2020**	**99.9**
Proposed IG+R-KNN	**2020**	**94.0**

5. Conclusion

In this paper, we presented a NIDS technique based on IG + Ranker method for IDS. This paper employed IG+R for feature selection stage, since this stage is known to be the most critical both in effort and time. The features were eliminated one at a time with lower rank from the bottom of the ranking and we observe the weight put by the ranker algorithm. The IG+R filter approach was carried out to remove both redundant and irrelevant features in the NSLKDD dataset. The feature selection with IG+R improves the performance of NB, SVM and KNN classifiers. The experimental results revealed that the three proposed IG+IR-NB, IG+IR-SVM and IG+IR-KNN outperformed existing techniques in terms of detection accuracy. In future, we planned to introduce a wrapper feature selection approach and compared the results with the proposed study.

References

[1] Z. Chiba, N. Abghour, K. Moussaid, A. El omri, and M. Rida, "Novel framework based on genetic algorithm and simulated annealing algorithm for optimization of BP neural network applied to network IDS," *ACM Int. Conf. Proceeding Ser.*, 2018, doi: 10.1145/3286606.3286805.

[2] A. A. Aburomman, M. Bin, and I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," vol. 38, pp. 360–372, 2016.

[3] C. Wang, R. Xu, S. Lee, and C. Lee, "Network Intrusion Detection Using Equality Constrained-Optimization-Based Extreme Learning Machines" *Knowledge-BasedSystems*,2018,doi: 10.1016/j.knosys.2018.02.015.

[4] S. Agrawal and J. Agrawal, "Survey on Anomaly Detection using Data Mining Techniques," *Procedia - Procedia Comput. Sci.*, vol. 60, pp. 708–713, 2015, doi: 10.1016/j.procs.2015.08.220.

[5] Ravi Kiran Varma P., Valli Kumari V., Srinivas Kumar S. (2018). "A Survey of Feature Selection Techniques in Intrusion Detection System": A Soft Computing Perspective. In: Pattnaik P., Rautaray S., Das H., Nayak J. (eds) Progress in Computing, Analytics and Networking. Advances in Intelligent Systems and Computing, vol.710. Springer, Singapore

[6] E. Ariafar and R. Kiani, "Intrusion detection system using an optimized framework based on datamining techniques," *2017 IEEE 4th Int. Conf. Knowledge-Based Eng. Innov. KBEI 2017*, vol. 2018-Janua, pp. 0785–0791, 2018, doi: 10.1109/KBEI.2017.8324903.

[7] A. Sultana and M. A. Jabbar, "Intelligent network intrusion detection system using data mining techniques," *Proc. 2016 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2016*, pp. 329–333, 2017, doi: 10.1109/ICATCCT.2016.7912017.

[8] H. Gharaee and H. Hosseinvand, "A new feature selection IDS based on genetic algorithm and SVM," *2016 8th Int. Symp. Telecommun. IST 2016*, pp. 139–144, 2017, doi: 10.1109/ISTEL.2016.7881798.

[9] Y. B. Reddy and R. Guha, "Intrusion detection using data mining techniques," *Proc. IASTED Int. Conf. Appl. Informatics*, pp. 26–30, 2004.

[10] S. Aljawarneh, M. Aldwairi, and M. Bani, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *J. Comput. Sci.*, vol. 25, pp. 152–160, 2018, doi: 10.1016/j.jocs.2017.03.006.

[11] M. Panda, A. Abraham, and M. R. Patra, "A hybrid intelligent approach for network intrusion detection," *Procedia Eng.*, vol. 30, no. 2011, pp. 1–9, 2012, doi: 10.1016/j.proeng.2012.01.827.

[12] M. Ektefa, S. Memar, F. Sidi, and L. S. Affendey, "Intrusion detection using data mining techniques," *Proc. - 2010 Int. Conf. Inf. Retr. Knowl. Manag. Explor. Invis. World, CAMP'10*, pp. 200–203, 2010, doi: 10.1109/INFRKM.2010.5466919.

[13] B. M. Aslahi-Shahri *et al.*, "A hybrid method consisting of GA and SVM for intrusion detection system," *Neural Comput. Appl.*, vol. 27, no. 6, pp. 1669–1676, 2016, doi: 10.1007/s00521-015-1964-2.

[14] M. S. R. and S. B. Xavier, "No TitleA hybrid intrusion detection system based on c5. 0 decision tree and one-class svm," *Int. J. Curr. Eng. Technol.*, vol. 5, no. 3, pp. 2001–2007, 2015.

[15] D. Perez, M. A. Astor, D. P. Abreu, and E. Scalise, "Intrusion detection in computer networks using hybrid machine learning techniques," *2017 43rd Lat. Am. Comput. Conf. CLEI 2017*, vol. 2017-Janua, pp. 1–10, 2017, doi: 10.1109/CLEI.2017.8226392.

[16] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *IEEE Symp. Comput. Intell. Secur. Def. Appl. CISDA 2009*, no. June 2014, 2009, doi: 10.1109/CISDA.2009.5356528.

[17] A. Özgür and H. Erdem, "A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015," *PeerJ*, vol. 4, pp. 0–21, 2016, doi: 10.7287/peerj.preprints.1954v1.

[18] W. Lee and S. J. Stolfo, *A Framework for Constructing Features and Models for Intrusion Detection Systems*, ACM Transactions on Information and System Secuirty, Vol. 3, no. 4. 2000.

[19] T. Z. Win, N. Saing, and M. Kham, "Information Gain Measured Feature Selection to Reduce High Dimensional Data," Seventeenth International Conference on Computer Applications (ICCA 2019) pp. 68–73.

[20] A. I. Pratiwi and Adiwijaya, "On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis," *Appl. Comput. Intell. Soft Comput.*, vol. 2018, 2018, doi: 10.1155/2018/1407817.

[21] Robert M. Gray, Entropy and Information Theory, Second Edition, Springer Science & Business media, 2011.

[22] J. Ding and L. Fu, "A Hybrid Feature Selection Algorithm Based on Information Gain and Sequential Forward Floating Search," *J. Intell. Comput.*, vol. 9, no. 3, p. 93, 2018, doi: 10.6025/jic/2018/9/3/93-101.

[23] B. Y. Ong, S. W. Goh, and C. Xu, "Sparsity adjusted information gain for feature selection in sentiment analysis," *Proc. - 2015 IEEE Int. Conf. Big Data, IEEE Big Data 2015*, pp. 2122–2128, 2015, doi: 10.1109/BigData.2015.7363995.

[24] J. Xu and H. Jiang, "An Improved Information Gain Feature Selection Algorithm for SVM Text Classifier," *Proc. - 2015 Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discov. CyberC 2015*, pp. 273–276, 2015, doi: 10.1109/CyberC.2015.53.

[25] Y. Zhang, X. Ren, and J. Zhang, "Intrusion detection method based on information gain and ReliefF feature selection," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2019-July, no. July, pp. 1–5, 2019, doi: 10.1109/IJCNN.2019.8851756.

[26] K. Suresh and R. Dillibabu, "Designing a Machine Learning Based Software Risk Assessment Model Using Naïve Bayes Algorithm," *TAGA J.*, vol. 14, pp. 3141–3147, 2018.

[27] I. D. Dinov and I. D. Dinov, *Probabilistic Learning: Classification Using Naive Bayes*. 2018.

[28] L. Li *et al.*, "A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset," *Genomics*, vol. 85, no. 1, pp. 16–23, 2005, doi: 10.1016/j.ygeno.2004.09.007.

[29] N. Cristianini and J. Shawe-Taylor: An Introduction to Support Vector Machines, Cambridge University Press (2000) "KJ00001304477.pdf.".

[30] W. Feng, J. Sun, L. Zhang, C. Cao, and Q. Yang, "A support vector machine based naive Bayes algorithm for spam filtering," *2016 IEEE 35th Int. Perform. Comput. Commun. Conf. IPCCC 2016*, 2017, doi: 10.1109/PCCC.2016.7820655.

[31] M. S. Neath, R.C., & Johnson, *Discrimination and Classification*. in International Encyclopedia of Education (Third Edition) 2010.

[32] G. Serpen and E. Aghaei, "Host-based misuse intrusion detection using PCA feature extraction and kNN classification algorithms," *Intell. Data Anal.*, vol. 22, no. 5, pp. 1101–1114, 2018, doi: 10.3233/IDA-173493.

[33] N. Moustafa and J. Slay, "A hybrid feature selection for network intrusion detection systems: Central points," pp. 5–13, 2017, doi: 10.4225/75/57a84d4fbefbb.

[34] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," *Proc. - IEEE Symp. Comput. Commun.*, no. Iscc, pp. 204–207, 2017, doi: 10.1109/ISCC.2017.8024530.

[35] Y. R. Somnay *et al.*, "Improving diagnostic recognition of primary hyperparathyroidism with machine learning," *Surg. (United States)*, vol. 161, no. 4, pp. 1113–1121, 2017, doi: 10.1016/j.surg.2016.09.044.

[36] S. Tandon, S. Tripathi, P. Saraswat, and C. Dabas, "Bitcoin Price Forecasting using LSTM and 10-Fold Cross validation," *2019 Int. Conf. Signal Process. Commun. ICSC 2019*, pp. 323–328, 2019, doi: 10.1109/ICSC45622.2019.8938251.

[37] C. Yin, Y. Zhu, J. Fei, and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," *IEEE Access*, vol. 5, pp. 21954–21961, 2017, doi: 10.1109/ACCESS.2017.2762418.

[38] K. R. S. A. Althubiti, E. M. Jones, "LSTM for Anomaly-Based Network Intrusion Detection," in *28th International Telecommunication Networks and Applications Conference (ITNAC), Sydney, NSW, Australia*, 2019, p. pp.1-3.

[39] X. C. T. Ma, Y. Yu, F. Wang, Q. Zhang, *A Hybrid Methodologies for Intrusion Detection Based Deep Neural Network with Support Vector Machine and Clustering Technique*, In: N. Yen. Springer, Singapore, 2018.

[40] S. C. R. Sharma, "An Enhanced Approach to Fuzzy C-means Clustering for Anomaly Detection," in *Proceedings of First International Conference on Smart System, Innovations and Computing, Smart Innovation, Systems and Technologies*, 2018, pp. 623–636.

[41] T. Omrani, A. Dallali, B. C. Rhaimi, and J. Fattahi, "Fusion of ANN and SVM classifiers for network attack detection," *2017 18th Int. Conf. Sci. Tech. Autom. Control Comput. Eng. STA 2017 - Proc.*, vol. 2018-Janua, pp. 374–377, 2018, doi: 10.1109/STA.2017.8314974.

[42] Sravanthi Godala, Rama Prasad V. Vaddella, "A Study on Intrusion Detection System in Wireless Sensor Networks", International Journal of Communication Networks and Information Security (IJCNIS), Vol. 12, No. 1, pp. 127-141, April 2020.