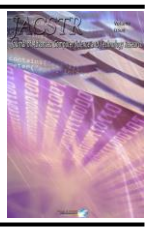

NSL-KDD dataset Classification Using Five Classification Methods and Three Feature Selection Strategies

Jamal H. Assi[1*], Ahmed T. Sadiq[2]

[1,2]Department of Computer Science, University of Technology, Baghdad, Iraq

[1*]jamalabomohammed@yahoo.com, [2]drahmaed_tark@yahoo.com

ISSN: 2231-8852

ABSTRACT

In this paper, five primary classification methods with three feature selection strategies have been implemented to classify the network attacks using NSL-KDD dataset. These methods are (J48 decision tree, Support Vector Machine (SVM), Decision Table (DT), Bayesian Network and Back Propagation Neural Network). The feature selection strategies are (Correlation base feature selection(CFS), Information Gain (IG) and Decision Table). Several experiments have been implemented to obtain good results using the training and testing NSL-KDD within general attack (Normal and Anomaly). These were carried out using four attack types: Denial of Service attack (DOS), User to Root attack (U2R), Remote to Local attack (R2L) and Probing attack. J48 classification method with information gain feature selection gives the best results (80.3%) using testing dataset and (93.9%) as an accuracy training dataset.

Keywords: classification methods, feature selection, NSL-KDD, Network attack

1. Introduction

Computer network attacks is a set of wicked activities to damage, refuse, degenerate or destroy information and service inside the computer networks. Now the world of computing is faced with the growing the chance of unintended downtime through to various attacks and security violation. In this environment of uncertain which is full of unauthorized access and wicked threats, those communities around the world which are the best at maintaining the stability of their services and retaining their computing power, enjoy a significant competitive advantage (Ali et al., 2010). Intrusion detection has turned into a major research field; many important network security mechanisms have been developed to prevent network attacks (Dhruba and Jagal, 2014). Network downtime results in financial losses and more harm to the trustiness of commercial activity, particularly ISPs (Internet Service Providers). Minimizing or possibly eliminating the unintended downtime of the system set up, ensures continuity of the computing services. Reducing unexpected

and unforeseen downtime can be done by characterizing, arrange and defending against misuse, attacks, and weakness points.

"Intrusion detection is the process of identifying and (possibly) responding to wicked activities targeted at computing and network resources." Any hardware or software computerization that monitors, detects or responds to events appear in a network or on a host computer is treated closely connected to the intrusion detection approach (Ali et al., 2010). This paper presents a review of comparing between machine learning classification methods applied to NSL-KDD training data set then using some algorithms for feature selection to decrease the dimensionality of the NSL-KDD data set then using the same classification methods and compare the results of different feature selection methods. The arrangement of this paper is as follows: Section 2 described related works, and Section 3 reviewed some machine learning classification algorithms. Sections 4, 5 and 6 looked into feature selection methods, a comparison between these algorithms based on its accuracy and runtime and the discussion of our results respectively. Finally, Section 7 gave the conclusion.

2. Related Works

In Shilpa et al. (2010), the number of computer resources, memory, and CPU are reduced by using a hybrid algorithm which is the Principle Component Analysis Neural Network Algorithm (PCANNA) used to discover attacks. Trained neural network that is used to identify the different kinds of new attacks, and features using the PCA transform are decreased. Enhanced version of NSL-KDD was utilized for testing and comparing of dataset; this has more advantages than KDD-CUP99. Analyzing dataset depending on NSL-KDD explains how the proposed model gives a more powerful and preferable representation. This analysis can decrease the resulted data features to 80.4% of data reduction. The percentages of the time reduction range between 70% (for time testing) and 40% (for time training).

In Dhanabal and Shantharajah (2015), the different classification algorithm (J48, SVM, and Naïve Bayes) has been used and analyzed by utilizing the NSL-KDD dataset. These algorithms are used to discover the abnormality in the networks packets. Moreover, NSL-KDD dataset is used to construe the protocols' connection obtained from the usually used network protocol stack through the intruders' attack that creates abnormal network traffic.

In Himadri et al. (2013), after experimenting with more than twenty most extensively used classification, ten algorithms are selected: J48, BayesNet, Logistic, SGD, IBK, JRip, PART, Random Forest, Random Tree and REPTree. The comparison of these ten classification algorithms depends on their performance metrics to discover the most appropriate and available algorithm the experimental results reveal that the Random Forest is the best one.

In Revathi and Malathi (2013), focus on a detailed study on NSL- KDD dataset that contains only selected record. Five classification algorithms (J48, Naïve Bayes, CART, Random Forest and SVM) were tested, the experimental results show that the SVM and Random Forest are the best methods.

In Hee-Su et al. (2013), focusing on feature selection or reduction leads to the elimination of some features because they are pointless and redundant which may cause corruption in the efficiency of the IDS. The goal of this research is to recognize significantly chosen input features in constructing IDS that is computationally robust and active. At this the evaluation of the performance of criterion feature selection approaches; Correlation-based Feature Selection (CFS), Information Gain (IG) and Gain Ratio (GR). They suggest a feature selection method using feature mediocre of overall and every class. The efficient classifier decision tree algorithm is applied for evaluating feature reduction method, then make a comparison between the proposed method and others. The Gain Ratio gives the best results.

Mukherjee and Sharma (2012) identify significantly decreased input features in constructing IDS that is computationally robust and active. The performance of the three standard feature selection methods was discussed. They proposed Feature Vitality Based Reduction Method to detect significantly reduced input features and applied one of the efficient classifiers Naive Bayes on reduced datasets for ID. Experimental results reveal that the selected features are decreased in their numbers to give us a better performance to design IDS that is characterized as solid and active NIDS.

In Dewa and Maglaras (2016), gives an insight of the existing Intrusion Detection Systems (IDS) along with their basic principles. Furthermore, it discusses how the data mining with its core feature (knowledge discovery) can help to create a data mining based on IDS. The resulted data mining may demonstrate more solid behavior comparing with traditional IDS and accomplish a higher accuracy to instruction's unique types.

In Shen et al. (2012), Several AI techniques like neural networks and fuzzy logic are applied in ID. The outcomes are diverse. The ID precision is the major concentrate for (IDS). The goal of the most research is to enhance the ID precision. On the one hand, they suggest an artificial immune system (AIS) based NID scheme and defined on the other hand an optimize feature opting using Rough set (RS) theory. The design of the selected algorithm addresses the complexity issue and tests the scheme using KDD CUP99 dataset version.

3. Classification Algorithms

The target of classification is to construct a model or classifier from something being classified so as to classify prior hidden objects as much accuracy as possible, rely on the information accessible on classes and the kind of classification. The result of a classifier may be given in different patterns, for instance in the pattern of decision trees or rules. The classification precision of most actual data mining algorithms requires being enhanced, as it is extremely hard to notice different novel attacks because the invaders unceasingly change its attack patterns (Chauhan, 2013). To resolve the data and classification of network attacks, some of the different classification algorithms like Bayesian Network, SVM, J48, MLP and Decision Table will be illustrating as below:

A) BayesNet

Is broadly used in classification, its build on the Bayes assumption it used conditional probability to calculate each node to construct a Bayesian Network. Bayes Net "is a directed acyclic graph (DAG) whose nodes are labeled by random variables" (Nilsson, 1998). Bayes Net N is a set of three (V, A, P) where:

1. V is a set of variables.
2. A is a set of arcs, which together with V to shape a directed acyclic graph G = (V, A).
3. P= {P (V|∏v): v ϵ V} where ∏v means the set of parents of v where P is a set of conditional probabilities of the whole variables given their corresponding parents.

This mechanism mostly used for IDS in a collection with statistical schemes. This step leads to many advantages (Nilsson, 1998):

1. The ability of encoding connections among variables and of forecast actions.
2. The ability to combine both ex- knowledge, and data.

There are some of the disadvantages such as (Nilsson, 1998):

1. Their outcomes are same to those came from threshold-based systems.
2. It required high calculation efforts.

B) SVM (Support Vector Machine)

SVMs is a mechanism appropriate for binary classification functions, which is associated to and includes elements of nonparametric enforce statistics, Neural Net, and machine learning. SVM "is a hopeful nonlinear, nonparametric classification technique, which previously displayed respectable outcomes in the medical diagnostics, optical character recognition, electric load forecasting and other fields" (Auria and Moro, 2008). SVM aims at discovering the function of the finest classification that differentiates members in training data of the two classes. The notion of the 'best' classification function can be measured and understood geometrically. For a linearly separable dataset, a linear classification function coincides to a dividing hyperplane $f(X)$ that passes through the center of the two classes, dividing the two. This moment a function is specified, new data instance X_n could classify by artlessly testing the sign of the function $f(X_n)$; X_n belongs to the positive class if $f(X_n) > 0$ (Wu et al., 2007).

So that there are many linear hyperplanes, Support Vector Machine is the best function found to maximize margin between two classes. The margin can be defined as "the amount of space, or separation between the two classes as specified by the hyperplane. Geometrically, the margin corresponds to the shortest distance between the closest data points to a point on the hyperplane" (Auria and Moro, 2008). The reason behind SVM insisting on finding the maximum-margin hyperplanes is that it present the best classification performance (e.g. Accuracy) on the training data and the best generalization ability (Wu et al., 2007).

C) J48 Classifier

J48 classifier algorithms can be described as a statistical classification that used to compare and create, using the notion of information entropy, a decision tree from a set of the training dataset.

These algorithms usually use the top down construction as a basic technique as an attempt to induce the decision tree for classification (Chauhan et al., 2013). J48 classifier algorithms are called a simple C4.5 decision tree for classifications. This decision tree is considered the most appropriate supervised classification technique that includes the simplest and fastest steps, classification and learning, which can be applied to any domain. Through the process of building any tree, J48 algorithms ignore all the missing values i.e. The values of any item can be predicted base on records on what is known about the features values (Patil and Sherekar, 2013).

D) Neural Networks

Neural Networks (NN) are considered the most important machine in learning techniques that used for classification, clustering …etc. It is an effort to figure machines that will simulate brain events to enable us to learn. NN typically learns by examples. For example, if NN is provided with adequate examples, it must be able to achieve classification and even realize novel drifts or shapes in data. Neural networks have many types, and most important one is Multi-level perceptron (MLP).

MLP "is an important type of NN, it is composed of multi-layers (input, output and hidden)." Each layer contains some nodes. According to the neural network, nodes can be classified into input nodes, hidden nodes, and output nodes. These three nodes are connected; the output nodes are the result of the connection of hidden and input layers. Back Propagation Algorithm is one of the popular NN algorithm that consists of four main steps:

1. "Feedforward computation."
2. "Backpropagation to the output layer."
3. "Backpropagation to the hidden layer."
4. "Weight updates."

This algorithm (back propagation) stops when the value of the error function becomes sufficiently small (Dmitrienko et al., 2014).

E) Decision Table

To search for the space of feature subsets effectively, we must transform the problem into a state space search and use the 'best- first search' to find the heuristical search (Xu et al., 1998; Kohavi, 1995). Features subset are the results of utilizing the 'best-first search.' The 'best-first search' includes operators that may add or delete features. The initial feature can either be a set of all features or an empty set. By doing so, the researcher tries to find the best optimal features. The researcher's selection of the decision table algorithm is to reduce the dimensionality of the dataset that evaluates feature subsets (Shen et al., 2012).

4. Feature Selection Methods

Feature selection is one of great significance the high dimension data makes testing and training of general classification methods difficult. In this paper, 3 feature selection methods have been

implemented which are Correlation based, Information Gain and Decision Table (mentioned above).

A) Correlation-Based Feature Selection (CFS)

CFS assess and grade feature subsets alternatively individual features. It chooses the set of features that are highly correlated with the class but with lower intercorrelation. CFS multi-heuristic seeks strategies such as 'hill climbing' and 'best first' that usually applied to search the feature subsets space in possible time. CFS first count a feature correlation and a matrix of feature class from the data training. Then, CFS searches the feature subset space utilizing the 'best first' (Mukherjee and Sharma, 2012). Using CFS helps in increase the performance of the machine learning (Hall, 1999).

B) Information Gain (IG)

The IG calculated features by measuring their 'information gain' on the class. It discretizes numeric attributes first using MDL based discretization method. Let C be set consisting of c data samples with m distinct classes. The training dataset c_i contains a sample of class I. Expected information needed to classify a given sample is calculated by (Hall, 1999):

$$I(c_1, c_2, \ldots . . c_n) = -\sum_{i=1}^{m} \frac{c_i}{c} log_2 \left(\frac{c_i}{c}\right) \ldots \ldots \ldots \ldots \ldots \ldots (1).$$

Where $\frac{c_i}{c}$ is the probability that an arbitrary sample belongs to class c_i. Let feature F has n distinct values $\{f_1, f_2, \ldots, f_n\}$ which can divide the training set into v subsets $\{C_1, C_2, \ldots, C_v\}$ where C_i is the subset which has the value f_i for feature F. The entropy of the feature F is given by

$$E(F) = \sum_{j=1}^{v} \frac{|c_i|}{|c|} I(c_i) \ldots \ldots \ldots \ldots \ldots \ldots (2)$$

Information gain for F can be calculated as (Mukherjee and Sharma, 2012):

$$Gain(F) = I(C) - E(F) \ldots \ldots \ldots \ldots \ldots \ldots (3)$$

C) NSL-KDD

The NSL-KDD data set is an enhanced version of the KDD cup99 data set. The inherent deficiency in the KDD cup99 dataset has been uncovered. Various statistical analyses have influenced the detection efficiency of many IDS modeled by researchers. NSL-KDD consist of vital records of the complete KDD dataset. NSL-KDD has the following features:

1. "Unnecessary records are removed to permit the classifiers to produce fair results."
2. "An adequate number of records is accessible by training and testing dataset, which is sensible reasonable and enables to perform tests on the full set."
3. "From each solid level group, the number of specific records is conversely genealogical to the records percentages in the original KDD dataset."

There are 41 attributes assigned to detect various features, in each record. The last attribute specifies the pattern, either as normal or anomaly (Dhanabal and Shantharajah, 2015).

The number of records in the training data set is 125972, and the number of records in the testing dataset is 22544. An attack establishes a connection between a source IP to a target IP address during a certain attack and sends data to attack the target (Dhanabal and Shantharajah, 2015):

1. Denial of Service Attack (DoS): "the attacker in this category, make some memory or coupling resources too busy to deny legitimate users access to their machine or manage legitimate requests." for instance: SYN flood and death's Ping.
2. User to Root Attack (U2R): "Attacker of this category access networks using a normal user account that sometimes exploited some debilitated as an attempt to gain root access to the system."
3. Remote to Local Attack (R2L): "In this category, attackers can send a specific packet to a specific machine over a network without knowing the account of that machine. Thus, this process exploits some debilitated that enable attackers to gain local access to the machine."
4. Probing Attack: "This category tries to find the system debilitated that assigned to attack system by gathering all the possible information concerning the network. For example, Port scanning (Dhanabal and Shantharajah, 2015).

5. Comparison Results for Intrusion Detection using Machine Learning Classification Methods

We often need to compare many different classification methods on the NSL-KDD dataset to obtain the right one to use. The results are performed using full training NSL-KDD dataset. Initially, we manage result on five different classifiers (BayesNet, SVM, J48, MLP and Decision Table). Table 1 shows the accuracy ratio and processing time for the above five classification methods within general 2 class (Normal and Anomaly) for training NSL-KDD dataset.

Table 1: Experimental Results of NSL-KDD Training Dataset in General

Method	NSL-KDD Accuracy	Time(sec.)
J48	99.846 %	30.09
BayesNet	90.6091 %	7.99
SVM	99.9667 %	54074.63
Decision Table	98.0432 %	124.66
Back-Propagation (MLP)	99.2379 %	99.3578

Table 2 shows the accuracy ratio and processing time for the above five classification methods within general five classes (Normal, DoS, Probe, U2R and R2L) i.e. one normal and four attack types for training NSL-KDD dataset.

Table 2: Experimental results of NSL-KDD Training Dataset for 4 Types of Attacks

Method	NSL-KDD Accuracy	Time(sec)
J48	99.9286 %	26.83
BayesNet	96.2571 %	3.41
SVM	99.8349 %	5228.94
Decision Table	97.5558 %	0.95
Back-Propagation (MLP)	99.3578 %	9.46

Table 3 shows the accuracy ratio and processing time for the above five classification methods using three feature selection methods within general five classes (Normal, DoS, Probe, U2R and R2L) i.e. one normal and four attack types for training NSL-KDD dataset.

Table 3: Experimental Results of NSL-KDD Training Dataset for 4 Types of Attacks Using Some Feature Selection Methods

Classification Method	Feature Selection Method	No. of Selected Att.	NSL-KDD Accuracy	Time
J48	Best first search	8	99.7698 %	4.69 sec.
		11	99.8404 %	4.33 sec.
		18	99.7095 %	7.83sec.
	Correlation	8	93.5636 %	5.54 sec.
		16	99.6253 %	8.18 sec.
		20	99.8746 %	12.51sec.
	Info Gain	8	80.5085 %	2.98sec.
		13	87.2881 %	6.66 sec.
		17	78.8136 %	10.18sec.
SVM	Best first search	8	72.3873 %	320.26 sec.
		11	79.661 %	5741.07 sec.
		18	78.8136 %	1084.9 sec.
	Correlation	8	75 %	2591.3 sec
		16	76.5152 %	6482.81sec
		20	77.9661 %	41742.27sec.
	Info Gain	8	76.2712 %	1.38 sec.
		13	77.9661 %	20579.24sec.
		17	76.2712 %	9729.52 sec.
	Best first search	8	68.5415 %	10.77 sec.
		11	79.661%	20.83 sec.
		18	48.3051%	33.91sec.
	Correlation	8	70.4545 %	0.14 sec.
		16	71.2121 %	27.37sec.

Decision Table		20	77.9661 %	40.11 sec.
	Info Gain	8	78.8136 %	15.4sec.
		13	75.4237 %	29.09 sec.
		17	75.4237 %	51.32sec.
Bayes Net	Best first search	8	96.5492 %	0.78 sec.
		11	98.4941 %	1.13 sec.
		18	93.4732 %	1.91 sec.
	Correlation	8	87.5218 %	1.64 sec.
		16	93.4541 %	1.69 sec.
		20	96.2563 %	3.38 sec.
	Info Gain	8	97.0946 %	1.16 sec.
		13	96.0777 %	1.89 sec.
		17	96.9922 %	2.55 sec.
Back Propagation	Best first search	8	86.8256 %	16.9 sec.
		11	88.0878 %	84.11sec.
		18	88.6999 %	151.98 sec.
	Correlation	8	86.1445 %	42.17 sec.
		16	89.4985 %	58.49 sec.
		20	88.5681 %	106.45 sec.
	Info Gain	8	88.6308 %	63.96 sec.
		13	88.9388 %	84.26sec.
		17	88.9944 %	102.03 sec.

Table 4: Experimental Results of NSL-KDD testing Dataset for 4 Types of Attacks using Some Feature Selection Methods

Classification Method	**Feature Selection Method**	**No. of Selected Att.**	**NSL-KDD Accuracy**	**Time**
J48	Best first search	8	72.96 %	7.09 seconds
		11	72.8803 %	5.34 seconds
		18	74.2536 %	10.68 seconds
	Correlation	8	68.3973 %	7.33 seconds
		16	73.6733 %	10.95 seconds
		20	78.1829 %	16.63 seconds
	Info Gain	8	78.0588 %	4.17 seconds
		13	80.9693 %	8.78 seconds
		17	74.1605 %	12.1 seconds
		8	72.424 %	228.31 seconds

SVM	Best first search	11	74.4042 %	181.78 seconds
		18	77.1286 %	72.4 seconds
	Correlation	8	68.003 %	105.79 seconds
		16	71.6134 %	9440.29 seconds
		20	73.4828 %	107.7 seconds
	Info Gain	8	74.6345 %	94.84 seconds
		13	71.7197 %	389.18 seconds
		17	71.5868 %	232.75 seconds
Decision Table	Best first search	8	72.96 %	7.09 seconds
		11	68.5833 %	20.64 seconds
		18	71.4627 %	32.5 seconds
	Correlation	8	64.406 %	8 seconds
		16	68.4903 %	36.15 seconds
		20	71.0729 %	47.21 seconds
	Info Gain	8	71.352 %	17.3 seconds
		13	67.0373 %	25.62 seconds
		17	66.674 %	51.16 seconds
Back Propagation	Best first search	8	68.0296 %	183.19 seconds
		11	72.681 %	3497.34 seconds
		18	73.8859 %	8224 seconds
	Correlation	8	68.8624 %	402.68 seconds
		16	69.0883 %	543.78 seconds
		20	72.2646 %	5518.78 seconds
	Info Gain	8	71.0508 %	1.26 seconds
		13	74.6833 %	0.66 seconds
		17	73.7397 %	0.71 seconds
BayesNet	Best first search	8	72.9911 %	0.86 seconds
		11	72.6411 %	1.78 seconds
		18	71.4627 %	32.5 seconds
	Correlation	8	65.5765 %	1.89 seconds
		16	70.497 %	2.76 seconds
		20	75.6135 %	0.42 seconds
	Info Gain	8	71.0508 %	1.26 seconds
		13	72.6322 %	0.13 seconds
		17	72.1051 %	0.11 seconds

By seeing the results and observations it appears that the weakness of the attack type R2L and if we excluded this type of attack the results as shown in table 5

Table 5: Experimental Results of NSL-KDD testing Dataset for 3 Types of Attacks (without R2L)

Classification Method	Feature Selection Method	No. of Selected Att.	NSL-KDD Accuracy	Time
J48	Best first search	8	83.4426 %	5.64 seconds
		11	85.5981 %	5.04 seconds
		18	82.6598 %	11.41 seconds
	Correlation	8	78.2522 %	7.84 seconds
		16	83.5138 %	8.65 seconds
		20	85.0389 %	16.57 seconds
	Info Gain	8	86.8283 %	4.82 seconds
		13	89.9243 %	10.17 seconds
		17	83.7629 %	11.9 seconds
SVM	Best first search	8	82.9597 %	11516.15 seconds
		11	83.6612 %	7149.86 seconds
		18	85.8167 %	1137.91 seconds
	Correlation	8	77.9472 %	7942.7 seconds
		16	82.004 %	8848.16 seconds
		20	83.3715 %	6362.29 seconds
	Info Gain	8	85.4456 %	4459.39 seconds
		13	82.0701 %	26720.61 seconds
		17	82.0548 %	14607.14 seconds
Decision Table	Best first search	8	79.5537 %	41.05 seconds
		11	79.5537 %	17.29 seconds
		18	77.8608 %	31.41 seconds
	Correlation	8	73.9871 %	9.64 seconds
		16	78.542 %	27.47 seconds
		20	82.8529 %	31.13 seconds
	Info Gain	8	81.8565 %	13.89 seconds
		13	77.1084 %	23.83 seconds
		17	76.6001 %	50.69 seconds
	Best first search	8	77.7795 %	153.76 seconds
		11	81.4804 %	4729.76 seconds
		18	84.6932 %	4755.09 seconds

Back Propagation	Correlation	8	78.8725 %	359.06 seconds
		16	79.9044 %	595.47 seconds
		20	81.8718 %	4113.47 seconds
	Info Gain	8	83.7273 %	4132.52 seconds
		13	84.0476 %	8491.75 seconds
		17	79.8333 %	6421.57 seconds
BayesNet	Best first search	8	83.5494 %	0.98 seconds
		11	83.402 %	1.38 seconds
		18	76.5747 %	0.11 seconds
	Correlation	8	75.868 %	0.23 seconds
		16	79.8892 %	0.09 seconds
		20	81.1245 %	3.36 seconds
	Info Gain	8	80.6263 %	1.38 seconds
		13	81.3329 %	2.7 seconds
		17	80.667 %	4.47 seconds

6. Discussion

The goal of this paper is to find the best classification algorithm that achieves the best performance on NSL-KDD dataset. The purpose is to analyze the NSL-KDD dataset and notice the performance of different classification algorithms. In Table 1, the training dataset in general where the class attribute is either normal or anomaly we see that the best accuracy among the five classification algorithms is in SVM it's (99.9667 %) but the time take to build the model is very high (54074.63 sec.) compared with others algorithm. However, in J48 algorithm the accuracy is (99.846%) and the time is 30.09 sec, in Table 2, training dataset using 4 types of attack (Dos, Probe, U2R,R2L) we notice that J48 algorithm is the best because the accuracy is (99.9286%) and the time is 26.83 sec. SVM is (99.8349%) but the time is 5228.94 sec. in Table 3, the training dataset with feature selection that use best first search, correlation, and information gain to reduce the number of features and choose the important features that directly affects the class attribute. in Table 3, we notice that the J48 classification algorithm with correlation method that reduces the features to 21 is the best performance (99.8746%) and the time is 12.51 sec.

7. Conclusions

The main idea was to obtain a good rate of accuracy and time required to build the classification model and reduce the false negative in intrusion detection. In future, we can combine between two algorithms to increase the performance and reduce the false negative. In the last table, we introduce four methods of machine learning classifiers with three methods of feature selection and show the results which are the correctly classified and incorrectly classified and the time that spends to build

the classifiers model. The feature selection methods are a correlation, information gain, and best first search to choose the relevant subset of features and ignore irrelevant features to gain possible best accuracy the number of attributes to enhanced the performance of the machine learning methods. The best classification methods as we saw above in general and four attack type and attack type without the R2L are J48 with feature selection Information Gain (80.9) accuracy and without R2L is (89.92) accuracy.

REFERENCES

1- Ghorbani Ali A., Lu Wei and Tavallaee Mahbod (2010). Network Intrusion Detection and Prevention Concepts and Techniques, Springer New York.

2- Dhruba Kumar Bhattacharyya, Jugal Kumar Kalita (2014) "Network Anomaly Detection a Machine Learning Perspective", CRC Press Tayler & Frances group, LLC.

3- Shilpa Lakhina, Sini Joseph and Bhupendra Verma (2010) "Feature Reduction using PCA for effective Anomaly-Based Intrusion detection on NSL-KDD", International Journal of Engineering Science and Technology Vol 2(6) pp1790-1799.

4- L. Dhanabal, and S. P. Shantharajah (2015) "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, Vol 4, Issue 6, pp.

5- Himadri Chauhan, Vipin Kumar, Sumit Pundir and Emmanuel S. Pilli (2013) "A Comparative Study of Classification Techniques for Intrusion Detection" International Symposium on Computational and Business Intelligence pp40-43.

6- S. Revathi, Dr. A. Malathi (2013) "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection and Technology, IJERT Vol. 2 Issue 12 pp1848-1853.

7- Hee-su Chae, Byung-oh Jo, Sang-Hyun Choi, Twae-kyung Park (2013) "Feature Selection For Intrusion Detection using NSL-KDD", Recent Advances in Computer Science, pp184-187.

8- Dr. Saurabh Mukherjee, Neelam Sharma (2012) "Intrusion Detection Using Naïve Bayes Classifier with Feature Reduction", Procedia Technology, Vol 4 (2012), pp119-128.

9- Zibusiso Dewa, Leandros Maglaras (2016) "Data Mining and Intrusion Detection Systems", International Journal of Advanced Computer Science and Applications, Vol 7 No 1, pp61-71.

10- Junyuan Shen, Jidong Wang, Hao Ai (2012) "An Improved Artificial Immune System Based Network Intrusion Detection by Using Rough Set", Communications and Networks, (2012)4, pp41-47.

11- Nils J. Nilsson (1998). Artificial Intelligence: A New Synthesis, Morgan Kaufmann Publishers, Inc., USA.

12- Laura Auria and Rouslan A. Moro (2008) "Support Vector Machines as a Techniques for Solvency Analysis", DIW Berlin Discussion Paper No. 811. http://dx.doi.org/10.2139/ssrn.1424949.

13- Xingdong Wu, Vipin Kumar, et. al. (2007) "Top 10 Algorithms in Data Mining" Knowledge Information System (2008)14, pp1-37
14- Tina R. Patil, S. S. Sherekar (2013) "Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification", International Journal of Computer Science and Applications, Vol. 6, No. 2, pp256-261.
15- Dmitrienko V. D., A. Yu. Zakovorotnyi, S. yu. Leonov and I. P. Khavina (2014) "Neural Network Art: Solving Problems with Multiple Solutions and New Teaching Algorithm", Open Neurol Journal (2014)8, pp15-21.
16- Lei Xu, Pingfan Yan and Tong Chang (1998), "Best First Strategy for Feature Selection", in 9th International Conference on Pattern Recognition, Rome, Italy. 14-17 November 1988.
17- Ron Kohavi, (1995) "The Power of Decision Tables", in Proceedings of 8th European Conference on Machine Learning, pp174-189
18- Mark A. Hall (1999) "Correlation Based Feature Selection for Machine Learning, Doctor of Philosophy Thesis, University of Waikato, Hamilton, New Zealand.