Combination of Multi Classification Algorithms for Intrusion Detection System

Dr. Maiwan Bahjat Abdulrazaq, Azar Abid Salih
Computer Science Department, Faculty of Science, University of Zakho

Abstract— Classification is one of the common tasks that are involved in data mining to build models for the prediction of future data. It performs its task by different classifier algorithms. This paper provides an approach based on information gain to determine the most distinguishing subset features of each attack class and combine multi classification algorithms which includes (Decision Tree J48, k nearest Neighbor and Naïve Bays). These classifiers are used for the task of detecting intrusions and comparing their relative performances. The goal of this work is to analyze the performance and accuracy of classification algorithms in order to identify the most efficient algorithm for each attack class, and then build accurate intrusion detection system. The proposed model has been applied on KDD Cup 99 data set using 60% of them for training and 40% for testing. These experimental results show that multiple classifiers work better than a single classifier. Also, multiple classifiers are more accurate and have abilities of distinguishing among the different attacks and normal connections effectively.

Index Terms— Classification Algorithms, Information Gain, Feature Selection, Data preprocessing, Intrusion Detection System, KDD Cup 99, Decision tree.

1 Introduction

Recently with a large number of internet users and different types of threats, the computer network security is becoming more important to defense information and detecting attacks. Thus, the Intrusion Detection System (IDS) is the best way possible to protect systems from anonymous attacks [1].

Intrusion is any attempt that threatens the integrity, confidentiality, availability of a resource, or to override the security mechanisms of a computer system or network [2]. The IDS can be defined as a process in network security for monitoring and analyzing events occurring in computer and network traffic system to detect signs of security problem and identify if there is any attack and helping in the protection of the system [3].
Generally, there are two techniques in analyzing the events using IDS: Misuse-(signature) based detection and anomaly based detection. In the misuse based IDS, a database of known attack signatures are maintained. Attacks are detected by comparing them with data collection unit and data stored in the database attack signatures. If a match occurs, then attack signal gets generated that they detect only the attacks for which they are trained to detect.

While anomaly based IDS detects attacks by observing deviations from the normal activities of the system by some history of monitoring systems, for example, earlier behavior or some previously defined profile of that system. The system matches the current profile with the previous profile to see if there is any significant deviation. Then, that activity is notified as an attack. The primary advantage of anomaly detection is the ability to detect novel and unknown attacks [4].

The IDS is a technique that provides security for both computers and networks. There are basically two types of intrusion detection systems [5]:
-Host-based IDS: It analyzes and monitors user activities and the operating system events on an individual computer that serve as hosts which are the internals of a computing system to detect the unauthorized intruders. Such as including audit record of operation system, system logs, application programs information.
-Network-based IDS: It analyzes and monitors network traffic from all computers. These systems collect information from the network itself and examine the traffic packet in real time. It capable of monitoring to detect remote attacks such as: Dos attacks, port scan.

2 Related works

This section discusses some of the techniques that used for designing and developing IDS. There are many works in the literature that performed it.

In 2012 Chandolikar and Nandavavdekar [6], proposed two data mining classification algorithms: Bayes net and J48 algorithm. Then , they analyzed and compared the results found that J48 learning algorithm was performing better than Bayes net in terms of better accuracy and lower error rate.In 2013, K. Kumar, et al [7], used different feature selection algorithms used such as Information Gain, Gain Ratio, One R, Relief etc. Combining the features of the best algorithms whose performance is better by comparing the results with each other using Decision Tree J48 classifier .They proved that the dimension reduction of data and select the most relevant features in intrusion detection can reduce training time, and increase classification accuracy. In 2013, Jayshri R. Patel [8], applied the decision tree classification algorithms such as C4.5 and CART, Random forest and Rep tree for intrusion detection. They used the information gain technique for the feature selection and decision tree for classification. They selected 15 features among 41 features for classification. The experiments showed that the performance of Random Forest is better as it correctly identifies more number of instances than the others and the decision tree classifier is able to improve the classification results for ranked intrusion detection. In 2014, V.Kosamkar and S.Chaudhari [9], proposed two hybrid algorithms for developing the intrusion detection system. C4.5 Decision Tree and Support Vector Machine (SVM) were tested with benchmark NSL- KDD data set. Correlation-Based Feature Selection (CFS) algorithm is used for feature selection.

3- BACKGROUND THEORY

3.1 Intrusion Detection System (IDS)

IDS plays important role in network security and it is one of the key technologies to guarantee the systems security. The goal of Intrusion Detection is to identify or detect wide variety type of attacks. The functions of intrusion detection system are [10]:

1- Analyzing and monitoring network traffic to classify whether the record is an attack or normal.
2- Analyzing system configurations and vulnerabilities.
3- Detect existing gaps in protection systems.
4- Ability to recognize and record all kinds of threats that occur in the network.
5- Identify the mistakes of the protection system administrator and abnormal activities

3.2 KDD CUP 99 DATA Set Description

A Since 1999, KDD Cup 99DataSet has been the most widely used data set for the evaluation of anomaly detection methods. Additionally, it is built based on the data captured in DARPA'98 IDS evaluation program which was prepared and managed by Lincoln Laboratory at MIT and it was used in the (Third International Knowledge Discovery and Data Mining tools competition), Which was held to evaluate the results of intrusion detection research [11].The KDD training dataset consists of approximately 4,900,000 single connection vectors.The feature or attribute 42 include target class labeled as either normal or attack as shown in Table 1.

Table 1: Whole KDD 99 Intrusion Detection Dataset

Dataset	Normal	Dos	Probing	R2L	U2R	Total
KDD	972,780	3,883,370	41,102	1126	52	4,898,430
10% KDD	97,278	391,458	4107	1126	52	494,021

The dataset used in the present study is a smaller subset (10% of the original training set), that contains 494,021. Every record in the dataset has 41 features or attributes including target class. The KDD 99 datasets are divided into two parts: the training dataset and the testing dataset.
The classes in KDD99 dataset can be categorized into five main classes (one normal class and four main intrusion classes: probe, DOS, U2R, and R2L) which are divided into 22 different attacks as shown in Table 2.

1) Normal: the normal connections are generated by simulated daily user behavior such as downloading files, visiting web pages [11].
2) Denial of Service Attack (DoS): is an attack in which the attacker makes some computing or memory resource too busy or complete to handle legitimate requests, or denies legitimate users access to computer system ,for example ping of death ,land, mail bomb, syn flood[11].
3) Probing Attack (Prob): is an attack in which the hacker scans a machine or a networking device in order to determine weaknesses or vulnerabilities that may later be exploited so as to compromise the system, for example: Port scanning [12].

4) User to Root (U2R) Attack: In this type of attack a local user on a machine is able to exploit some vulnerability to obtain privileges normally reserved for the root users, for example: buffer overflow attack [13].
5) Remote to Local (R2L) Attack: In this type of attack an attacker who has the ability to send packets to a machine over a network but does not have an account on a victim machine exploits some vulnerability to obtain privileges local access to the machine and modifies the data for example: password guessing[13].

Table2: Attacks in KDD 99 Training dataset

Class number	Four main Attack Classes	22 Attacks Classes	Sum of records	No. of attacks
1	Normal		87831	1
2	Denial of Service (DoS)	back, land, neptune, pod, smurf, teardrop	54572	6
3	Probing (Prob)	ipsweep, nmap, portsweep, satan	2131	4
4	Remote to User (R2L)	ftp_write, Guess_passwd, Imap, multihop, phf, spy,warezclient, warezmaster	999	8
5	User to Root (U2R)	buffer_overflow, perl, loadmodule, rootkit	52	4
Total	**5 class**		**145585**	**23**

4 Frame Work

There many stages depended to achieve the proposed system. The stages are started at preprocessing data it consists of data cleaning from the noise such as inconsistent, missing values and removing duplicate to handling model to get the best result for classifying, and then transform the features that has text forms to numeric forms. This is an important operation for machine learning classification. Additional that the data scaled within a small specified range (0 to 1) by using Min-Max normalization. After that this model can effectively select the most distinguish features for intrusion detection rate by using a reduced features set (Information Gain) with several kinds of classification algorithms including (Decision Tree j48, K Nearest Neighbor and Naïve Bayes) to identify which algorithm is better and get best results classification accuracy of each type of attack class.
The input data used in the system is the intrusion detection dataset called KDD'99 cup dataset. The classifier is trained 60% of the data set and theremainder for the testing set.

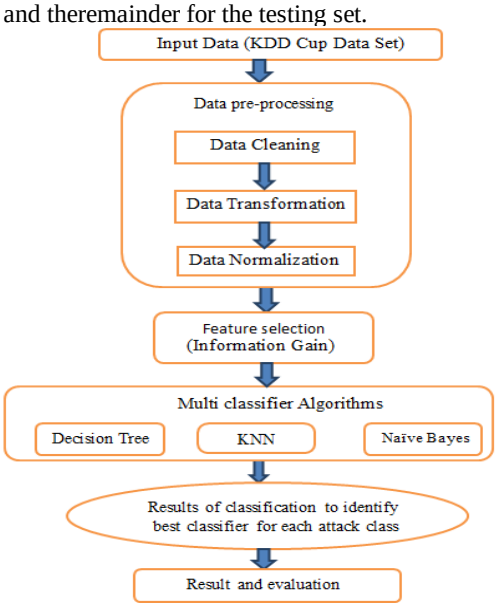

Fig.1. Flowchart of general structure of the proposed system

4.1 Data Pre-Processing

In real world data are generally dirty such as it contains errors, missing value, duplicate, outlier, incomplete, irrelevant and inconsistent data. The purpose of data preprocessing is to clean the noise data, extract features, and transforms the preliminary data into a format that will be more easily and effectively processed for the purpose of the user. This technique helps to improve the efficiency of the algorithm to classify the data correctly.

The data preprocessing involves three stages

1- Data cleaning: this stage is responsible for removing any records contains a missing value and inconsistent value. It also removes duplicate records in the data set.
2- Data transformation: after data cleaning the next stage of pre-processing is to transform or convert the features that have text forms to numeric form to be suitable input for classification algorithms.
3- Data normalization: as a step of data preprocessing when datasets are too large, attribute normalization is important to detection performance. The ranges of the feature are normalized by scaling it is value so that they fall within the small specified range 0 to 1). The method min- max normalization is applied.

Min-Max Normalization: performs a linear transformation on the original data values. Suppose that min and max are the minimum and maximum of feature x. mapping interval $[min_x, max_x]$ into a new interval to $[new_{min_A}, new_{max_A}]$.Consequently, every value v from the original interval will be mapped into value new_v using

$$v' = \frac{v - min_A}{max_A - min_A}(new_{max_A} - new_{min_A}) + new_{min_A} \quad \ldots(1)$$

Table 3: The result of pre processing

Class	Number of instances				
	KDD 10%	After removing missing value	After removing duplicated instances	Reduction instances	Percent all classes
Normal	97278	97277	87831	9.7%	60.33%
Dos	391458	391458	54572	86.1%	37.48%
Prob	4107	4107	2131	48.1%	1.46%
R2L	1126	1126	999	11.3%	0.69%
U2R	52	52	52	0.0%	0.04%
Total No. of instances	494021	494020	145585	70.5% reduction	100%
				29.5% corrected instances	

5 Feature Selection: is a process to select the most relevant feature set by removing irrelevant or redundant features. Only subsets of original features are selected.
The Objectives of feature selection techniques are [14]:
-Reduce the dimensionality of feature space to avoid in curse of dimensionality.
-Speed of learning algorithms.
-Reduce large amount of data
- Less memory storage.
- Increasing the training (learning model) performance.
- Enhance Predictive accuracy by decreasing overfitting.
-Sampling the data more efficiently and improving the data quality.

6 Information Gain (IG)

In this method, the features are filtered to create the most influential feature subset before the start of the learning process. This algorithm using rankers method on features and evaluate the feature by ranking them from most important to least important.
Entropy is commonly used in the information theory measure of the impurity in the collection of training data.

The feature has higher entropy is more information content.
Given a collection of instances S
Proportion of S belonging to class i

$$\text{Entropy}(S) = \sum_{k=0}^{n} -Pi \ \text{Log2}\ (Pi) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Information gain of an attribute A is defined as:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{|Sv|}{|S|} \text{Entropy}\ (S_v) \quad \ldots\ldots\ldots\ldots(3)$$

A: Is the set of all possible values for attribute A
SV: is the subset of S for which A has value
Only the highest attributes ranking is used by the classification algorithms to classify the input dataset to either normal or attack as shown in Table 4.

Table 4: Attribute ranking by Information Gain

Class	No. features	Rank Search (Info Gain)
Dos	20	30,29,3,5,23,4,34,35,33,6,38,39,25,26,12,32,36,37,31,24
Prob	20	5,3,6,35,29,37,23,27,33,12,40,4,30,34,25,38,36,41,2,28
R2L	11	5,3,6,33,10,36,37,22,24,23,1
U2R	8	3,14,10,33,1,13,17,32

7 CLASSIFICATION

Classification is a supervised learning technique. It is one of the important tasks of data mining predictive modeling that classify the data item into predefined class label [15].
In the classification the following techniques are mainly used like Decision Trees (j48), Bayesian (Naive Bayes) and Lazy (K-Nearest Neighbor).Before classification, data preprocessing techniques and feature selection are applied this techniques improve the efficiency of the algorithm to classify the data correctly.

7.1 Classification Algorithm

Differentiation classification algorithms have been used for the performance evaluation, below are listed.

7.1.1 Decision Tree (J48)

In classification, there are a number of tree algorithms the most popular tree classifier is J48 decision tree which was primarily designed as the enhanced version of C4.5 that builds decision trees by using the concept of information entropy. It uses the fact that each attribute of the data can be used to make a decision by splitting the data into smaller subsets which consist of nodes that form a rooted tree. The tree has three types of nodes. The first is a root node that has no incoming edges, the second one is Internal nodes (branches) and all other nodes are called leaves also known as terminal or decision nodes, each leaf node is assigned a class label, and other internal nodes contain attribute test conditions to separate records that have different characteristics [16].

For building decision tree, the pseudo code for building Decision Trees is written below

1. Check for base classes
2.For each attribute a find the information gain from splitting a
3. Let A is a best attribute with the highest information gain.
4. Create a decision node that splits the A
5. Recurs on the sub lists obtained by splitting a best and add those nodes as children's of nodes.

7.1.2 Naïve Bayes

Bayesian classifiers a simple approach based on the inferences of probabilistic graphical models. It is very easy to construct, not needing any complicated iterative parameter estimation schemes and also predicts the class label in the fastest time [17].

The Naive Bayes classifier is a supervised learning algorithm based applying Bayes' theorem. This method assumes that all the features are independent values of each other predictors. This assumption is called class conditional independence.

Using Bayesian theory

$$P(Ci|X) = \frac{P(X|Ci).P(Ci)}{P(X)} \quad \text{……..(4)}$$

$$P(Ci|X) = \frac{P(x1|Ci).P(x2|Ci)....P(xn|Ci).P(Ci)}{P(X)} \quad \text{……...(5)}$$

$$P(Ci|X) = P(x1|Ci).P(x2|Ci)....P(xn|Ci).P(Ci) \quad \text{…..(6)}$$

Let X: be the data record is a possible value in the session class

Ci: represent the category of classes

P(Ci |X):is the posterior probability of class.

$P(Ci)$: is the prior probability of the class.

$P(X|Ci)$: is the probability of class given predictor attribute.

P(X): is the prior probability of the predictor.

C1=Normal, C2=DOS, C3= Prob, C4=R2L, C5=U2R

7.1.3 k Nearest Neighbor (KNN)

It is one of supervised classification lazy learning which takes more time for predicting. It uses the distance weighting measures with the capability attributes of thesearch method and only valid for continuous variables.

K is the number of nearest neighbors.

Using Euclidean distance between two input vectors with m quantitative features where x = (x1,…., xm) and y = (y1,…., ym).

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \quad \text{……..…..(7)}$$

Let x be the point to be labeled. Find the point closest to x, let it be y.

KNN Procedure steps:

1- Store the KDD99 training data set with corresponding label.

2- K: Number of nearest neighbour.

3- Calculate the distance between (x', y') test and every example (x, y) training, select the minimum nearest to neighbour.

4- Repeat the procedure for all connection in testing dataset.

5- For End.

Disadvantages K Nearest Neighbor [18]:

1-large space of memory requirement to store the complete training

2-That lazy learning methods take more time are usually slower to classify data set.

8 Training Data Stage

It is a dataset that is used to build a classifier which is a process of learning something from instances in order to predict the class attributes of new coming unknown instances. Each instance in the training set contains one "target value" is a class label. The learning phase, where the training data is analyzed and classification rules are generated.

9 Testing Data stage

It is a dataset that is used to evaluate the learning algorithm. It is used to validate the learning algorithm. These learning algorithms when applied to training dataset makes that dataset learn to work according to that algorithm. Then to check if that algorithm is appropriate to be used with that dataset is validated by evaluating that dataset with the testing dataset. The testing phase is the classification, where test data is classified into classes according to the generated rules.

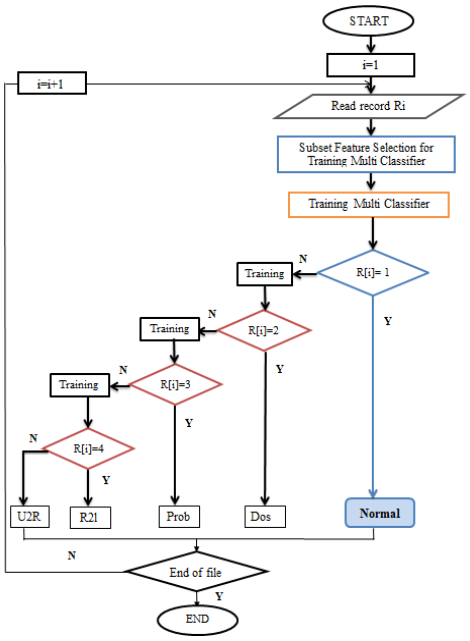

Fig.2. Flowchart of Training stage

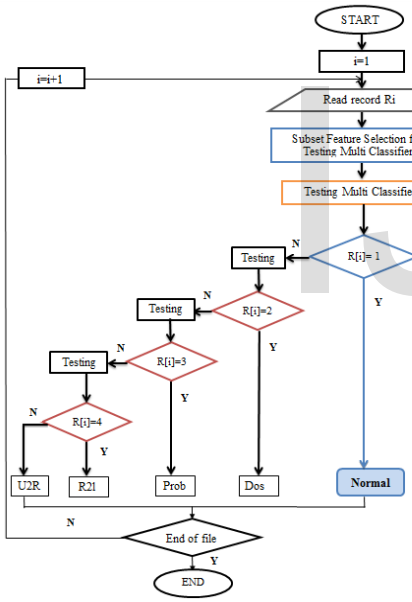

Fig.3. Flowchart of Testing stage

10 Experiments and Result

The performances metrics are calculated as follows:
True positive (TP): Attacks correctly classified as attacks
False positive (FP): Incorrectly classified normal data as attacks
True negative(TN): Normal correctly classified as normal
False negative (FN): Incorrectly classified attacks as a normal

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP+FN} = \frac{Correct\ Intrusions}{Intrusions} \quad \ldots\ldots\ldots (8)$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP+TN} = \frac{Normal\ as\ Intrusions}{Normal} \quad \ldots\ldots (9)$$

$$\text{True Negative Rate (TNR)} = \frac{TN}{TN+FP} = \frac{Correct\ Normal}{Normal} \quad \ldots\ldots\ldots (10)$$

$$\text{False Negative (FNR)} = \frac{FN}{FN+TP} = \frac{Intrusions\ as\ Normal}{Intrusions} \quad \ldots\ldots (11)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{Correct\ Classification}{All\ Instances} \quad \ldots\ldots\ldots\ldots (12)$$

$$\text{Error rate} = 1 - \text{Accuracy} \quad \ldots\ldots\ldots\ldots (13)$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{Correct\ Intrusions}{Instances\ Classified\ as\ intrusions} \quad \ldots\ldots\ldots\ldots (14)$$

Confusion Matrix: information about actual and predicted classifications done by a classification system. For two classes case can be represented as shown in table (5).

Table 5: Confusion Matrix

		Predicted Class	
	Activity	Attack	Normal
Class	Attack	TP	FN
	Normal	FP	TN

Table 6: Number of Records for Training 60% and Testing 40%

	Train	Ratio	Test	Ratio	sum	Ratio
Normal	52699	36.20	35132	24.13	87831	60.33
DOS	32743	22.49	21829	15	54572	37.49
Probing	1279	0.88	852	0.59	2131	1.47
R2L	599	0.41	400	0.27	999	0.68
U2R	31	0.02	21	0.01	52	0.03
Total	87351	60	58234	40	145585	100

10.1 Training results

Table 7: 60% Training Decision Tre

Parameter	Dos	Prob	R2L	U2R
True Positive	1.000	1.000	1.000	1.000
False Positive	0.000	0.007	0.026	0.193
Precision	1.000	1.000	1.000	1.000
Recall	1.000	1.000	1.000	1.000
F-Measure	1.000	1.000	1.000	1.000
Time to build model	9.45 sec	5.68 sec	1.6 sec	0.79 sec
Accuracy Correctly Classified Instances	99.99	99.98	99.96	99.99
Error rate Incorrectly Classified Instances	0.01	0.02	0.04	0.01

Table 8: 60% Training K Nearest Neighbor

Parameter	Dos	Prob	R2L	U2R
True Positive	1.000	1.000	1.000	1.000
False Positive	0.000	0.000	0.000	0.000
Precision	1.000	1.000	1.000	1.000
Recall	1.000	1.000	1.000	1.000
F-Measure	1.000	1.000	1.000	1.000
Time to build model	0.013sec	0.05 sec	0.02 sec	0.02 sec
Accuracy Correctly Classified Instances	100	100	100	100
Error rate Incorrectly Classified Instances	0	0	0	0

Table 9: 60% Training Naïve Bayes

Parameter	Dos	Prob	R2L	U2R
True Positive	0.975	0.952	0.981	0.992
False Positive	0.034	0.074	0.119	0.258
Precision	0.976	0.982	0.991	0.999
Recall	0.975	0.952	0.981	0.992
F-Measure	0.975	0.963	0.985	0.995
Time to build model	0.59 sec	0.41 sec	0.31 sec	0.84sec
Accuracy Correctly Classified Instances	97.54	95.21	98.13	99.15
Error rate Incorrectly Classified Instances	2.46	4.79	1.87	0.85

10.2 Testing results

For testing stage the number of records for each class (Normal= 53132, Dos = 21829 , Prob= 852 , R2l =400 , U2R= 21)

Table 10: 40% Testing Decision Tree

Classes Attacks	No. of Records	Incorrect Records Classified	Detection Rate
Dos	56961	169	99.70
Prob	35984	705	98.04
R2l	35532	10	99.97
U2R	35153	9	99.97

Table 11: 40% Testing K Nearest Neighbor

Classes Attacks	No. of Records	Incorrect Records Classified	Detection Rate
Dos	56961	9787	82.82
Prob	35984	1703	95.27
R2l	35532	6	99.98
U2R	35153	3	99.99

Table 12: 40% Testing Naïve Bayes

Classes Attack	No. of Records	Incorrect Records Classified	Detection Rate
Dos	56961	56	99.90
Prob	35984	8688	75.86
R2l	35532	4909	86.20
U2R	35153	4169	88.14

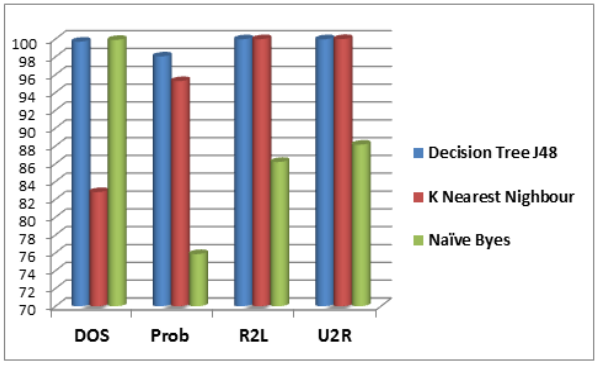

Fig.4. Comparison chart of accuracy Classifiers for each attack class. (Testing Results)

11 CONCLUSIONS

In this paper, the classification algorithms Decision Tree (j48), lazy (KNN), Bayesian (Naïve bays) are discussed and analyzed the classification accuracy of each type of class. The results show that applied preprocessing data and relevant feature selection using information gain for reducing features of the data set are very important to decrease the learning time of the algorithm and increase of the classification accuracy.

The accuracy of classification gets the best result when (Prob) class is used with Decision Tree classifier while (R2L and U2R) are used with K Nearest Nighbour classifier, and the (Dos) is used with Naïve Bayes. The combination among DT, KNN and NB may give better accuracy for the above classes. The Naïve Bayes classifier under-performs and gives less accuracy than any of the two other classifiers but it is a faster classifier for building a model and the KNN is slower as it takes more time to test a model for training classifier.

References

[1] ReyadhSh.Naoum, Wafa' S. Al-Sharafat, "Adaptive Framework for Network Intrusion Detection by Using Genetic-Based Machine Learning Algorithm", IJCSNS International Journal of Computer Science and Network Security, VOL.9No.4. April 2009.

[2] Megha Aggarwal, Amrita,"Performance Analysis of Different Feature-Selection Methods in Intrusion Detection", International journal of science & technology research volume 2, Issue 6, June 2013.

[3] Lata, InduKashyap, " Study and Analysis of Network based Intrusion Detection System", International Journal of Advanced Research in-Computer and Communication Engineering Vol. 2, Issue 5, May 2013

[4] Devikrishna K, Ramakrishna B, "An Artificial Neural Network based

Intrusion Detection System and Classification of Attacks ", Vol. 3, Issue 4, Jul-Aug 2013.

[5] G.V. Nadiammai, S.Krishnaveni, M. Hemalatha, "A Comprehensive Analysis and study in Intrusion Detection System using Data Mining Techniques", International Journal of Computer Applications, Volume 35– No.8, December 2011.

[6] N.S.CHANDOLIKAR, V.D.NANDAVADEKAR, "Comparative Analysis of two Algorithms for intrusion attack classification using KDD CUP DataSet", International Journal of Computer Science and Engineering (IJCSE) Vol.1, Issue 1 Aug 2012.

[7] Krishan Kumar, Gulshan Kumar, Yogesh Kumar, " Feature Selection Approach for Intrusion Detection System", International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE), Vol.2, No.5, 2013.

[8] JAYSHRI R. PATEL, " Performance Evaluation of Decision Tree Classifiers for Ranked Features of Intrusion Detection", NOV 12 TO OCT 13 | volum – 02, ISSUE – 02.

[9] Vaishali Kosamkar, Sangita S Chaudhari," Improved Intrusion Detection System using C4.5 Decision Tree and Support Vector Machine", International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014.

[10] Prabhdeep Kaur & Shevetavashisht, " Evaluation of Intrusion Detection Technique and Algorithms in Terms of Performance and Efficiency through Data Mining", International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) Vol. 3, Issue 2, Jun 2013.

[11] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani " A Detailed Analysis of the KDD CUP 99 Data Set", Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009).

[12] Swati Paliwal, Ravindra Gupta ," Denial-of-Service, Probing & Remote to User (R2L) Attack Detection using Genetic Algorithm", International Journal of Computer Applications ,Volume 60– No.19, December 2012.

[13] Rupali Datti, ShilpaLakhina,"Performance Comparison of Features Reduction Techniques for Intrusion Detection System", IJCST Vol. 3, Issue 1, Jan. - March 2012.

[14] Tanya Garg, AmandeepKaur, "A Review Paper on Data Mining Approach for Feature Selection for Network Intrusion Detection System", Volume 3, Issue 11, November 2013.

[15] Syeda FarhaShazmeen, Mirza Mustafa Ali Baig, M.Reena Pawar, "Performance Evaluation of Different Data Mining Classification Algorithm and Predictive Analysis", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 10, Issue 6 (May. - Jun. 2013).

[16]- Upendra,"An Efficient Feature Reduction Comparison of Machine Learning Algorithms for IntrusionDetection System", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) , Volume 2, Issue 1, January – February 2013.

[17]-MrutyunjayaPanda1, Manas RanjanPatra, "Evaluating Machine Learning Algorithms for Detecting Network Intrusions", International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009.

[18]- S. Vijayarani1, M. Muthulakshmi, "Comparative Analysis of Bayes and Lazy Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 8, August 2013.

Dr. Maiwan Bahjat Abdulrazzaq received his BSc at 1991 in Computer Science Dept. Department at Mosul University, and Master degree from Al Neealin University, Faculty of Computing Science & Information Technology, Computer Science Dept. at 2003, and PhD degrees from University of Zakho, Faculty of Science, Computer Science Dept. at 2013. He is lecturer in University of Zakho.

Azar Abid Salih, received his BSc at 2009 in Computer Science Dept. at University of Duhok.He is working now to get MSc degree in Computer Science. This paper is one of his research part productions working together with his supervisor Dr. Maiwan Bahjat Abdulrazaq.