Volume 6, Issue 9, September 2016 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Empirical Analysis of Classification Approaches for Feature Selection in Intrusion Detection

Rajinder Kaur[*]
MTech Scholar, Department of Computer Science & Engg.
SBSSTC, Ferozepur, India

Monika Sachdeva
Associate Prof. Department of Computer Science & Engg.
SBSSTC, Ferozepur, India

Gulshan Kumar
Assistant Prof. Department of Computer Application
SBSSTC, Ferozepur, India

***Abstract**— As the data is growing on the network day by day, the Intrusion detection has become the major research problem in the network security. Intrusion detection is a process of monitoring, detecting and analyzing the network traffic data to find out the security violations in the networks. The one of the essential challenge in Intrusion detection is to deal with data containing the huge amount of features, which represents the whole dataset. Feature Selection and Machine learning approaches facilitate to design the "Intrusion Detection Systems" which are classifying the network traffic data into intrusive traffic or normal traffic. Network traffic dataset has contained some of redundant and irrelevant features; the usage of such features in Intrusion detection can decrease the efficiency as well as increase the computational time for proficient response in the real time environment. In this work, have tried to reduce the redundant and irrelevant features by using different feature selection approaches and estimation has been done by using the top-seven classification algorithms namely Bayes Net, Naive Bayes, J48, Random Forest, OneR, PART and Decision Tree are selected on the basis of their speed and ability to handle large dataset. The performance of different classification models is calculated using 10-fold cross validation. The experiments and evaluation of these approaches is performed in WEKA data mining tool by using the benchmark KDD-99 dataset. Finally the empirical results indicate that the Naive Bayes and PART classifier outer perform best in detection rate and OneR classifier takes less time to train the model.*

***Keywords**— Feature Selection, Intrusion Detection, Naive bayes, OneR, WEKA, Classification Accuracy.*

I. INTRODUCTION

Due to big amount of data contains on the network, the data is contented vulnerable for the various forms of attacks i.e. rise in intrusions is increasing day by day. Intrusion detection is extremely significant to prevent the intruders to smash into or misuse the system. To protect against the various types computer viruses and network attacks a lot of techniques have been developed. From these techniques the Network Intrusion Detection (NID) has been taken as the most confident technique to defend from complex and active intrusion behaviours [1]. Intrusion Detection Systems (IDS) categorize the data into two categories as normal and intrusive. In IDS different classification algorithms have been proposed to devise an effective Intrusion Detection process. However the performance of intrusion detection system is depending on the performance of the classifier. So the selection of appropriate classifier assists to enhance the performance of the intrusion detection system.

1.1 Intrusion Detection

The Intrusion Detection System is a process of monitoring the actions and activities occurring in a computer system or the network system and analyzing the signs of possible incidents. An IDS was first launched in 1980 by James. P. Anderson [2] and then enhanced by D. Denning [3] in 1987. Intrusion Detection Systems (IDSs) play a critical role as a defence mechanism in the network security. IDS preserve the network security objectives like data integrity, confidentiality and availability of system from the attacks. Intrusion detection approaches can be divided into two categories as anomaly detection and misuse detection. Anomaly detection finds out attacks on basis of the patterns that diverge from the normal behaviour of system. Misuse detection determines the attacks on basis of the patterns that are extracted from the known intrusions. While the classification accuracy is the necessity of IDS, the extensibility and flexibility are also essential in the network computing surroundings. In network Intrusion Detection System a huge quantity of activity data is gathered from network and generating the large log files and unprocessed network traffic data, at which human assessment is impossible. So, these activity data have to be compressed into sophisticated events, called features and attributes. More than it, a set of features is attained and monitored by IDS to detect the intrusion attempts. Yet, there are some features with false relationships, hiding the principal process, and that may be also redundant and irrelevant. In this means, discarding these features and selecting an optimal features set that adequately express the network environment are important in order to accomplish the fast and useful response against the attack attempts, reduce the dimensionality of data and the computation time, and boost the precision of the IDS. In this way, Data mining approach feature selection is capable to efficiently remove patterns of intrusions for misuse detection, launch profiles of normal network actions for anomaly detection, and construct classifiers to detect attacks.

© 2016, IJARCSSE All Rights Reserved

1.2 Feature Selection

Feature Selection (FS) is one of the important data pre-processing tool used in the field of Data Mining. At present the data is increasing in terms of features and instances at the network that to reduce the processing time and to achieve the higher detection rate in form of accuracy results. Feature Selection process discard the irrelevant and redundant features from the network traffic dataset. Feature selection refers to approaches that are selecting the best subset of features from the original set features. FS process selects a subset of unique features according to particular criteria and frequently used dimensionality reduction approach for data mining [4]. In this research work, we have reduced the features from the Network Intrusion Detection Dataset, for which we have applied the selection based and ranking based feature selection approaches that are choosing the features according to their importance. To analyze the intrusions 41 features are considered in our experiments. The aim of this research is to reduce the number of unnecessary features so in order to intrusions can be identified in small period of time with best accuracy results. Compare the output of feature selection approaches using different classifiers in WEKA tool. The effectiveness of the different classifiers method is evaluated by carry out several experiments on benchmark KDDCup99 network intrusion dataset. There are numerous performance metrics to calculate performance of Feature Selection approaches such as True Positive (TP) rate, False Postive (FP) rate, Receiving Operating Curve (ROC) area, Kappa statistic, Classification accuracy and time to train the model.

The rest of this paper is organized as follows: Section 2 presents a background of the classifiers, including Bayes Net, Naive Bayes, J48, Random Forest, OneR, PART and Decision Tree. Section 3 describes Experimental Setup & Methodology including the NSL-KDD 99 Dataset and WEKA environment tool. Section 4 introduces the Result & Discussions with performance metrics for feature selection in IDS. Finally, Section 5 concludes the results.

II. CLASSIFICATION

Classifiers are the widely used supervised data mining approach known as machine learning or classification algorithms. The main goal of the classification is to build the model to classify the unseen objects accurately on basis of classified objects [5, 6]. The type of classification depends on the information contains on the classes such as decision trees or rules form. Classifiers are applied to classify the network traffic dataset as normal or intrusive traffic and if dataset is intrusive then it describe the category of intrusive such as probe, Dos, U2R or R2L. In this research work, there are 72 classifiers which are compatible of our opted dataset and we have evaluated the best seven classifiers with different performance metrics. The Classification process involves the following steps:

- Pre-processing the training data set.
- Identify class feature and classes.
- Identify useful features for the classification
- Learn a model by using training examples in the Training set.
- Use the learnt model to classify the unknown data samples.
- Evaluate the results by testing phase.

2.1 Bayes Net

The Bayes Net (BN) algorithm is most widely used for the classification purpose and it is network based framework for expressing and analysing forms involving the uncertainty [7]. Bayes Net classifier based on the Bayes theorem in which probability on each feature is evaluated from the Bayesian Network. From other knowledge – based classifiers the Bayes Net is different because uncertainty in BN is handled mathematically rigorous so it is efficient and simple. Bayesian networks execute the incremental learning from the causal relationships.

2.2 Naive Bayes

The simplified Bayesian probability model is used in Naive Bayes and it performs the classification task with more efficiency. In this, it is considered that possibility of single attributes never effect over the possibility of other attributes. Naive Bayes can be combined with some of attribute selection techniques to eliminate redundant and irrelevant data [7, 13]. It has numerous advantages including the capacity of encoding interdependencies among the variables and of the calculating events, also the capability to integrate both prior data and knowledge.

2.3 J48

J48 classifier is frequently referred to as a numerical classifier that generate a decision tree from the labelled data of training dataset by using the model of information entropy and it observe same results from opting a feature for dividing the data [8]. It is a rule based classifier. This algorithm utilized the greedy top-down production method to persuade the decision trees for the classification. The dividing procedure stops only if all the features in the subset belong to the same category.

2.4 Random Forest

Random forest is an ensemble classification algorithm and it's depend upon the decision tree algorithm and generating the output in the form of entity trees. The working algorithm of Random forest is a combination of the random selection of features set and bagging idea in order to construct a group of decision trees with prohibited variation. In a random forest, every node is divide using the best surrounded by a subset of predictors randomly selected at same node.

It handles the large number of variables without ignoring any variable in accurate manner. It optionally produces the additional pieces of information as a measure of importance of predictor variables and as a measure of the inside structure of data.

2.5 OneR

OneR known as “One Rule” is a straightforward classification algorithm that produces a one-level decision tree. OneR is competent to infer typically effortless, accurate and classification rules from a group of features. It creates one rule for all features in the training data, and then decides the rule with the least error rate since its ‘one rule’. To make a rule for a feature, most frequent class is determined for every feature. A rule is basically a group of feature values bound to their popularity class. It calculates the total errors occurred at the rules of every predictor.

2.6 PART

PART is a rule based classifier. It utilized the separate and conquers method to construct a C4.5 decision tree for all iterations that identifies the "BEST" for the rule. PART is an attribute based learner and it uses an entropic distance quantify.

2.7 Decision Tree

Decision Tree (DT) is the most widely well-known machine learning approach produced by Quinlan [9]. DT includes main three components nodes, arcs, and leaves. Each node splits further in the instance space and into two or more categories according to a particular discrete function at the input feature values. The root node is known as the test node and it has no inward edge. Many decision tree building algorithms engage a two – step process. Initial, a very big decision tree is developed. After that, to decrease large dimension and over fit the data, in the next step, the specified tree is trimmed. The trimmed decision tree have been used for the classification purposes and known as classification tree.

III. EXPERIMENTAL SETUP AND METHODOLOGY

3.1 WEKA Tool

WEKA stands for Waikato Environment for Knowledge Analysis, is a well-liked collection of machine learning software’s that are written in Java language and developed by the University of Waikato in New Zealand. WEKA tool is free available software. The WEKA environments contain a group of visualization tools, algorithms for predictive modelling and data analysis. The functionality in WEKA is very easy due to it has a user graphical interface and it is familiar as a landmark scheme in data mining and machine learning process. Algorithms can be applied directly over the dataset or may be prepared by personal Java code. WEKA provides the facility of 76 classification algorithms, 49 data pre-processing tools, 15 attribute evaluators and 10 search algorithms for purpose of feature selection. In Graphical User Interfaces contains "The Explorer", "The Experimenter" and "The Knowledge Flow" module. The WEKA tool stored the data in Attribute Relation File format (ARFF) file format. WEKA supports numerous standards of data mining tasks, data pre-processing tasks, classification, clustering, regression, visualization, and feature selection. It runs over any recent computing platform [10, 14].

3.2 Network Traffic Dataset

Since 1999, KDD99 are most widely used dataset for the assessment of anomaly detection methods. The data set is organized by [11], and it is developed on basis of captured data in DARPA98 IDS evaluation program. The dataset is used offline for the analysis and measure the performance of Intrusion detection system. The main aim is to provide a benchmark dataset for the researcher to develop the new approaches against the different network security threats like e-mail, tablet, web etc.In benchmark KDD99 dataset, there contains 494,021 instances from which 97,278 instances are considered as a normal and 396,744 are labelled with attacked of 22 different types that are classified into 4 main categories. The labelled connection vector consists of 41 features and has 1 attack type.Simulated attacks may fall down into one of the subsequent four categories: User to Root Attack (U2R), Remote to Local Attack (R2L), Denial of Service Attack (DOS), Probing Attack or Normal. In KDD99 training and testing dataset has large number of connection records for the classifiers. So, to reduce the non-uniformity in dataset, we have selected randomly maximum of 44,000 connection records of all attack type and used for the reason of training and testing of the classifiers. We randomly selected the 37,791 connection records for training dataset and 6753 connection records for the testing purpose as indicate in table I.

Table I. Description of Connection records used in Dataset

Class Type	# Training dataset Instance	# Testing dataset Instance
Normal	12533	1609
Dos	11656	1607
Probe	12555	1628
U2R	52	200
R2L	995	1719
Total Records	37791	6753

3.3 Research Methodology

The purpose of this research study is to analyze the effect classifiers on the detection accuracy with different number of features to classify the instances of the network traffic dataset. Maximum accuracy with less number of features is necessary to minimize the computational time of the intrusion detection system. In this study, we have been considered the seven classifiers to evaluate the performance of the selected features. An open source WEKA data mining tool is used to carry out this research. We have performed experiments on Intel i3 M 380 2.40 GHz processor with 6GB of RAM with Windows 7 operating systems. We performed 5-class classification of benchmark KDD- 99 dataset. We conducted group of experiments by using default parameters of WEKA applied classifiers. The research methodology implemented for this study is described as in following steps and in Fig. 1 .

3.3.1. *Pre-processing step:* In this step, symbolic form of features is converted into numeric form. The normalized form of features is used for the training and testing phases.

3.3.2 *Feature selection step*: In this step, Correlation based feature selection (CFS), Chi-squared feature selection, Filtered attribute evaluator and Gain ratio feature selection approaches are applied to discard the redundant and irrelevant features from the training and testing dataset.

3.3.3 *Classification step:* Classification steps have two phases namely as training phase and testing phase. In Training phase the classifier is learnt by the reduced training dataset and output of this phase is trained by using 10 cross validation. Whereas in Testing Phase, the trained model is given input of the Test dataset and revaluate the model to predict the class label.

3.3.4 *Performance metrics evaluation:* After the testing phase, the performance metric evaluation step computes the defined performance metrics such as True positive rate, False Positive rate, Kappa statistics, ROC area, Classification accuracy and training time. The values of these performance metrics lies in range (0-1).

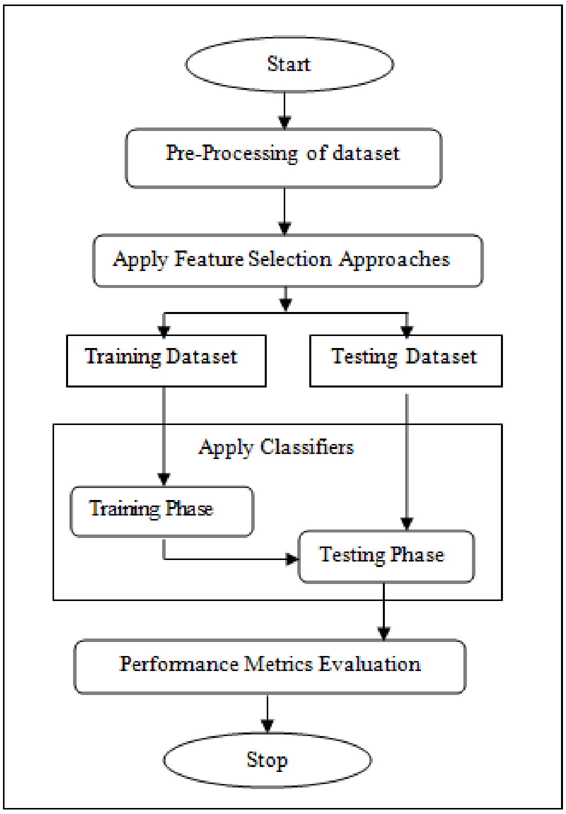
Start
Pre-Processing of dataset
Apply Feature Selection Approaches
Training Dataset
Testing Dataset
Apply Classifiers
Training Phase
Testing Phase
Performance Metrics Evaluation
Stop

Fig.1. Experiment Methodology

IV. RESULT & DISCUSSIONS

4.1 Performance Metrics

There are different parameters used to analysis the results of classification model. In our experiments we have analyzed the performance of different classifier with True positive (TP) rate, False positive (FP) rate, Kappa statistics, ROC area, Classification Accuracy and Training time. Table II indicates the confusion matrix that is used to calculate TP rate, FP rate and accuracy. Confusion matrix summarizes the number of instances calculate normal or abnormal by the classification model.

Table II. Confusion Matrix

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

© 2016, IJARCSSE All Rights Reserved

4.1.1 *True Negative*- TN determines the number of detected, normal instances which are normal in actuality.

4.1.2 *False Negative*- FN determines the number of detected, normal instances which are actually attacked.

4.1.3 *True Positive*- TP determines the number of detected attacks which are actually attacked. TP rate is defined as the ratio of number of classified attack connections and full amount of normal connections.

$$TP\ Rate = \frac{TP}{TP+FN}$$

4.1.4 *False Positive*- FP determines the number of detected attacks which are normal in actuality. FP rate is defined as the ratio of the number of misclassified normal connections and full amount of normal connections.

$$FP\ Rate = \frac{FP}{FP+TN}$$

4.1.5 *Accuracy*- To measure the performance of the classifier the classification accuracy (CA) is most required. It concludes the fraction of correctly classified Instances over the full amount of instances.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

4.1.6 *Receiving Operating Characteristic (ROC Area)*- ROC is applied to draw a curve between TP Rate and FP Rate and the area contained under the curve is known as AUC that gives the value of the ROC.

4.1.7 *Kappa Statistic- This* is a statistic which calculates the inter-rater contract for qualitative or categorical items. The value of the kappa statistic lies between 0 to 1 ranges. 0 means totally disagree and 1 means full agreement.

4.1.8 *Training time*- It is total time used by Classifier to construct the model on a given dataset. It is frequently calculated in seconds.

4.2 Result Analysis

Table III indicates the different features recommended by various feature selection approaches. In these experiments, we have analysed the various existing feature selection approaches with the use of different search methods. These feature selection approaches are further compared by using various performance metrics like TP Rate, FP Rate, ROC Area, Kappa Statistic, Classification Accuracy, and Training Time. Then picked the best subset of feature selection approaches scheduled on the basis of performance metrics. Existing FS that are employed in experiments are Cfs Subset Eval, Chi Squared Attribute Eval, Filtered Attribute Eval, and Gain Ratio with search methods Best First search and Ranker. These search methods seek for the set of all probable features in order to obtain a best subset of feature.

Table III. Features suggested by different feature selection approaches.

Feature Selection Approach	*No of feature selection*	*Selected Features*
CFS + Best First Search	10	3,5,6,12,14,26,30,34,37,38
Chi Squared Eval + Ranker	15	5,3,6,35,33,4,30,23,34,37,38, 29,25, 39,36
Gain Ratio + Ranker	18	12,26,39,6,22,14,25,4,5,30,38,11, 10,3,17,18,37,29
Filtered Attribute Eval + Ranker	20	5,3,35,30,4,6,33,23,29,34,38,25, 39,37,,26,36,12,32,24,31

The optimal selected subset of features is future employed at various classifiers. Table IV Show the performance of the different classifiers at full amount of dataset which includes the 41 features. Empirical results indicate that the full amount of dataset take more time to train the model and it has large dimensionality of data which gives less results in True Positive rate, Accuracy and in Kappa statistics. At 41 Features dataset PART Classifier outer perform best than other classifiers in rate of classification accuracy (53.4887%), True Positive rate (0.545), ROC (0.664) and Kappa Statistic (0.4978). Decision tree gives less false positive rate (0.133) and OneR best in Training time with amount of only 34.31 sec. These results give us an idea about the need of feature selection approaches in order to reduce the dataset.

Table IV. Comparative results of different Classifiers at 41 Features Dataset.

Classifier	*TP rate*	*FP rate*	*ROC area*	*Kappa Statistic*	*Accuracy (%)*	*Time Taken (Sec)*
Bayes Net	0.457	0.137	0.72	0.538	44.6754 %	46.3
Naive Bayes	0.542	0.143	0.727	0.4635	52.4831 %	43.98
J48	0.514	0.171	0.722	0.3935	51.3929 %	47.34
Random Forest	0.535	0.146	0.716	0.4631	52.4154 %	76.03
OneR	0.493	0.159	0.667	0.3332	49.2681 %	**34.31**
PART	**0.545**	0.146	**0.664**	**0.4978**	**53.4887 %**	51.73
Decision Tree	0.532	**0.133**	0.775	0.4111	52.1597 %	80.31

Empirical results indicate that the reduced data set features gives best result in detection rate. Table V show that correlation based feature selection with best first search method suggest the 10 features from 41 features. At reduced dataset PART, Random Forest, J48 and Decision Tree classifier provide the best results in accuracy, True positive rate, ROC area and Kappa Statistic. Whereas the OneR, Naïve Bayes and Bayes Net take less time to Train the model. Detection rate almost 15% increased at reduced dataset.

Table V. Comparative results of different Classifiers at 10 Features Dataset suggested by CFS feature selection approach.

Classifier	*TP rate*	*FP rate*	*ROC area*	*Kappa Statistic*	*Accuracy (%)*	*Time Taken(Sec)*
Bayes Net	0.589	0.127	0.896	0.4614	58.8792 %	3.62
Naive Bayes	0.52	0.14	0.892	0.3766	52.0035 %	3.16
J48	0.614	0.107	0.839	0.4996	61.3633 %	7.68
Random Forest	0.636	0.105	0.868	0.5268	63.5813 %	33.15
OneR	0.493	0.159	0.667	0.3332	49.2681 %	**3.14**
PART	**0.681**	**0.099**	**0.865**	**0.5809**	**68.0615 %**	13.19
Decision Tree	0.601	0.125	0.896	0.4202	60.0917 %	15.09

From Table VI, VII and VIII it is cleared that the firstly Naïve bayes classifiers performances best at 15, 18 and 20 subset of features. It gives best results in intrusion detection. It enhances the True Postive rate, ROC area, Kappa Statistic and also lowers the False positive rate up to 0.068. Naïve bayes classifier gives 66.8342% Accuracy at 15 features, 63. 6256% at 18 features and 73.1037% at subset of 20 features, which indicates that it is adequate classifier to detect the attacks in competent manner. After the Naïve bayes PART, J48 and Random Forest performs best in Intrusion detection. OneR, Naïve bayes and Bayes Net takes less time to train the model. Empirical results also show that at 20 subset of features classification accuracy rate enhance than all other subsets. It proves that Filtered attribute evaluator approach is best approach in feature selection.

Table VI. Comparative results of different Classifiers at 15 Features Dataset suggested by ChiSquared feature selection approach.

Classifier	*TP rate*	*FP rate*	*ROC area*	*Kappa Statistic*	*Accuracy (%)*	*Time Taken (Sec)*
Bayes Net	0.586	0.12	0.904	0.4585	58.6426 %	6.51
Naive Bayes	**0.668**	**0.096**	**0.912**	**0.5677**	**66.8342 %**	5.32
J48	0.649	0.099	0.809	0.5439	64.912 %	15.11
Random Forest	0.623	0.117	0.853	0.5054	62.3244 %	43.75
OneR	0.493	0.159	0.667	0.3332	51.8681 %	**5.24**
PART	0.648	0.109	0.796	0.538	64.8085 %	26.56
Decision Tree	0.594	0.127	0.875	0.4664	59.3819 %	12.48

Table VII. Comparative results of different Classifiers at 18 Features Dataset suggested by Gain Ratio selection approach.

Classifier	*TP rate*	*FP rate*	*ROC area*	*Kappa Statistic*	*Accuracy (%)*	*Time Taken (Sec)*
Bayes Net	0.591	0.124	0.905	0.4653	59.0566 %	6.96
Naive Bayes	**0.636**	**0.109**	**0.919**	**0.5243**	**63.6256 %**	6.4
J48	0.625	0.117	0.767	0.5085	62.5166 %	17.06
Random Forest	0.625	0.117	0.885	0.5087	62.5314 %	46.04
OneR	0.593	0.159	0.667	0.3332	59.2681 %	**6.04**
PART	0.627	0.116	0.835	0.5105	62.6793 %	28.56
Decision Tree	0.558	0.139	0.89	0.4202	55.8332 %	15.62

Table VIII. Comparative results of different Classifiers at 20 Features Dataset suggested by Filtered Attribute evaluator selection approach.

Classifier	*TP rate*	*FP rate*	*ROC area*	*Kappa Statistic*	*Accuracy (%)*	*Time Taken (Sec)*
Bayes Net	0.599	0.122	0.913	0.4758	59.9438 %	7.97
Naive Bayes	**0.731**	**0.068**	**0.935**	**0.6523**	**73.1037 %**	7.11
J48	0.617	0.115	0.803	0.4988	61.659 %	19.21
Random Forest	0.606	0.123	0.855	0.4826	60.5796 %	40.95
OneR	0.493	0.159	0.667	0.3332	49.2381 %	**6.72**
PART	0.62	0.119	0.856	0.501	61.9991 %	35.35
Decision Tree	0.577	0.133	0.884	0.4441	57.6815 %	16.11

Figure 2. the comparison analysis of accuracy rate and Figure 3. Shows comparison study of Training time with Bayes net, Naïve bayes, J48, Random Forest, OneR, PART, Decision Tree classifiers with different feature selection approaches. Results indicate that the Naïve bayes classifier and PART classifiers outer perform best than all other classifiers in Accuracy rate. Whereas, the OneR classifier is a best classifier in Training time.

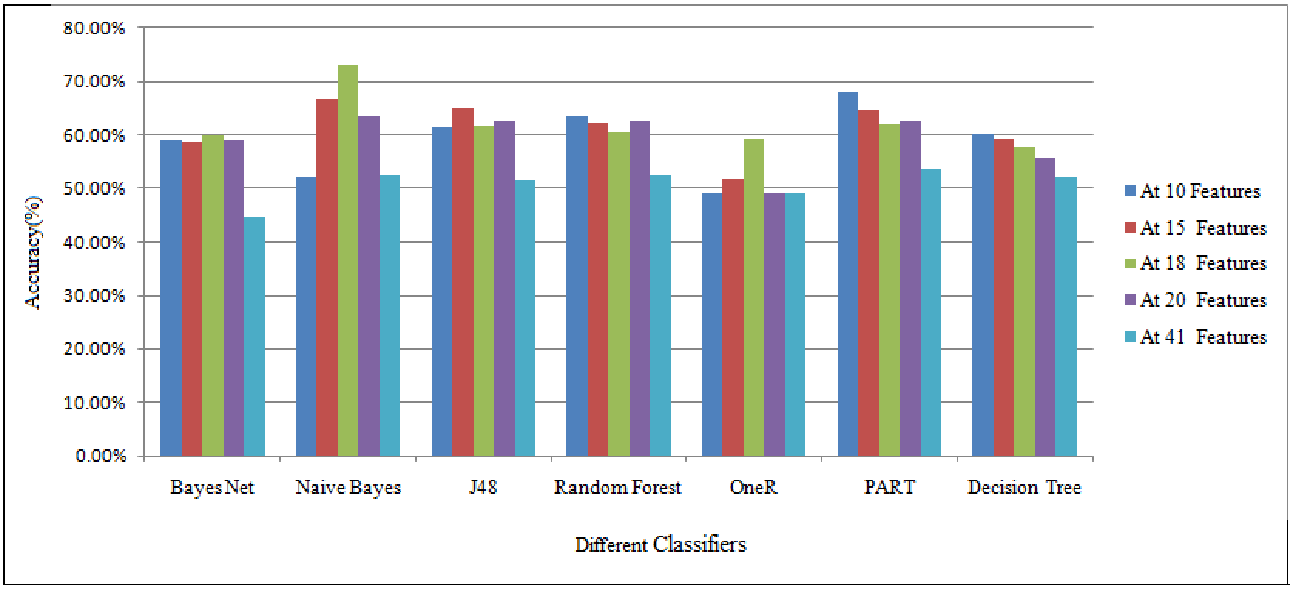

Fig. 2 Accuracy rate Comparison of different classifiers

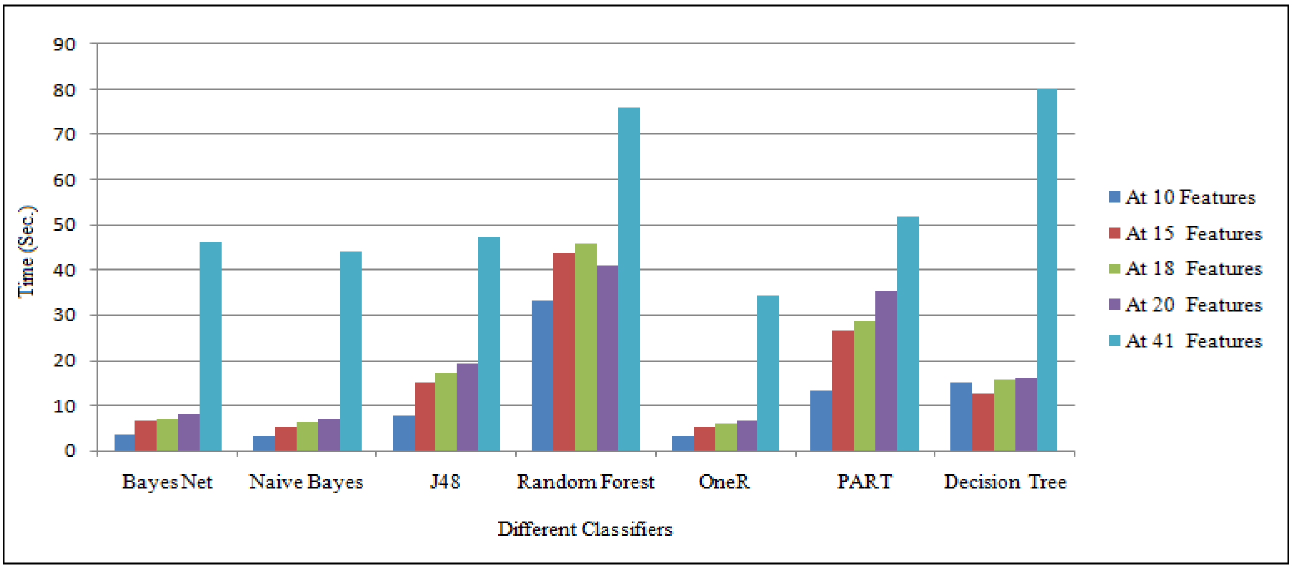

Fig. 3 Training Time Comparison of different classifiers

V. CONCLUSIONS

In this empirical research work, we performed a set of experiments of various supervised classifiers at benchmark KDD-99 dataset having 41 features. In order to discard the redundant and irrelevant features form large dataset the CFS, ChiSquared, Gain Ratio, Filtered attribute evaluator feature selection approaches are employed. The performance of Naive bayes, Bayes net, J48, Random Forest, OneR, PART and Decision tree classifiers are compared using various performance metrics such as TP rate, FP rate, classification accuracy, ROC Area, Kappa Statistic and Training Time. Main intention of this research work is to analyze the common supervised classifiers that are used in literature for the intrusion detection. Empirical results of experiments conclude that Naive bayes and PART classifiers are best and stable classifier for association concerned with the accurate classification of malicious traffic network dataset with maximum TP rate, ROC area, Accuracy and minimum FP rate. Next, it is also recommended that rule based J48, Random Forest and OneR classifiers can be utilized for the detection of various attack classes. In future work will focus on enhancing the results of intrusion detection by combing the various feature selection approaches with a set of classifiers.

REFERENCES

[1] R. Dash, Selection of the Best Classifier from Different Datasets Using WEKA, IJERT, Vol.2 Issue 3, March 2013.

[2] James P. Anderson, “Computer Security Threat Monitoring and Surveillance,” Technical Report, James P.Anderson Co.,Fort Washington, Pennsylvania, USA , pp.98–17, April 1980.

[3] Dorothy E. Denning,“An Intrusion Detection Model,” IEEE Transaction on Software Engineering (TSE), volume–13, No.2, pp.222–232,February 1987.

© 2016, IJARCSSE All Rights Reserved

[4] P. A. Est´evez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," Neural Networks, *IEEE Transactions on*, vol. 20, no. 2, pp. 189–201, 2009.

[5] K. Kumar, G. Kumar, Y. Kumar, Feature selection approach for intrusion detection system, International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE) 2 (5).

[6] F. Amiri, M.M.R. Yousefi, C. L. A. S. N. Y. (2011). Mutual information-based feature selection for intrusion detection systems. J. Netw. Comput. Applications, 34(4):1184-1199.

[7] P. Langley, W. Iba, K. Thompson, "An analysis of Bayesian classifiers*", Tenth National Conference on Artificial Intelligence*, 1992, pp. 223-228.

[8] G. Kalyani and A. J. Lakshmi, "Performance Assessment of Different Classification Techniques for Intrusion Detection," *IOSR Journal of Computer Engineering (IOSRJCE),* vol. 7, no. 5, pp. 25-29, 2012.

[9] Salzberg, S. L. (1994). Book review: C4.5: Programs for machine learning by J. Ross Quinlan. Machine Learning, 16, 235–240.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The WEKA data mining software: An update, SIGKDD Explorations 11 (1) (2009) 10-18.

[11] KDD, "Kddcup99 intrusion dataset." [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/

[12] G. Kumar, K. Kumar, and M. Sachdeva, "The use of artificial intelligence based techniques for intrusion detection: a review," *Artificial Intelligence Review,* vol. 34, no. 4, pp. 369–387, 2010.

[13] G. K. Ahuja, K. K. Saluja, and M. Sachdeva, "An empirical comparative analysis of feature reduction methods for intrusion detection," *International Journal of Information and Telecommunication Technology (ISSN: 0976-5972),* vol. 1, no. 1, 2010.

[14] "Waikato environment for knowledge analysis (weka) version 3.7". [Available online]: http://www.cs.waikato.ac.nz/ml/weka/, June, 2008.