Study and Comparison of Feature Selection Approaches for Intrusion Detection

Rajinder Kaur
MTech Scholar
Shaheed Bhagat Singh State Technical Campus
Ferozepur, Punjab, INDIA

Monika Sachdeva
Associate professor
Shaheed Bhagat Singh StateTechnical Campus
Ferozepur, Punjab, INDIA

Gulshan Kumar
Associate professor
Shaheed Bhagat Singh State Technical Campus
Ferozepur, Punjab, INDIA

ABSTRACT
At Present, it is very essential to establish a high level network security to make sure the more trusted and secure communication between various organizations. Network Security provides a platform to secure information channels from the huge amount of network attacks. Intrusion Detection System (IDS) is an estimable tool for the defense mechanism in computer networks. IDS focus on detecting of harmful network traffic that would exploit vulnerability in network system. Feature selection performs a necessary role in intrusion detection process. The dataset extracted in IDS contain a large number of features, in which some of irrelevant, redundant and noisy. These unnecessary features degrades the performance of the IDS. In order to discard irrelevant, redundant & noisy features in the experiment, have need to analyzed different feature selection approaches with various search methods. The pre-processed NSL-KDD dataset is used in experiments for evaluation purpose at WEKA 3.6.9 environment tool. By using Bayes Net and Naive Bayes Classifier classify the selected feature dataset. The comparison of all empirical results are done by using different performance metrics. The ultimate goal of work is to increase the overall accuracy of the detection process with minimal number of selected feature dataset and reduced training time.

Keywords
Intrusion Detection System (IDS), Feature Selection (FS) Approaches, Pre-Processing Dataset, Bayes Net, Naive Bayes, NSL-KDD

1. INTRODUCTION
In 1987 Dorothy E. Denning proposed intrusion detection as a technique to counter, the computer and networking attacks and misuses [1]. An intrusion detection system is needed in computer network systems as a defense mechanism against the malicious attacks. IDS implement the process of intrusion detection in network security system. Intrusion is defined as the group of activities & actions that attempt to break the security objectives like Integrity, Availability, Assurance, Confidentiality etc. [2]. The security policies of the system are disrupted by the various intrusions. Intruders basically are two types - legitimate users of the network and from outside of the network. Intrusion Detection is process of measuring, analyzing, and controlling the events occurring in a system. The clues of security problems are detected by intrusion detection. An Intrusion detection system (IDS) controls the network traffic, monitors for incredulous activities and alerts the network administrator. IDS are defined as a security approach which can detect, avoid and may be reacting to threats and computer attacks. It is a powerful & an efficient security system which implements the intrusion detection process & reports the intrusions precisely to the appropriate system authority.

Main Components of IDS- *Network to monitor* is a single host or the network host component that determines the instructions for the monitoring process. *Data Collection & Storage* unit responsible for data collection at distinct events occur in a system network and also transform them into proper form and to store at a particular disk. *Data Analysis & Processing unit* are intelligent and central part of the IDS work like a brain in human beings. It maintains the whole process to detect the huge amount of attacks. Feature selection approaches are used in this unit. A signal is generated while detecting any attack and sent to the network administrator. Further *Network security Administrator* control all the outputs and give a response by alert and alarm to Network to Monitor. (Refer Fig. 1).

IDS system can be classified into two categories on the basis of **Detection approaches and Depending on location**. By location IDS classify as *Host based IDS* and *Network based IDS*. Host based IDS (HIDS) estimate the information found lying on a single or several host systems which include the contents of the application system, Operating system & files. Network IDS proficient to access the network routers & instruct them to carry out various responsibilities. NIDS estimate the information captured from network infrastructure and analyzing the stream of packets that can be passed over the network. On the basis of Detection approaches IDS can be distinct into *Misuse based IDS* and *Anomaly based IDS*. Signature or Misuse based IDS performs the task of matching the activities of the user through stored signatures of well known attacks. When found any sign of attack, and then it indicates the signal. Whereas Behavior or Anomaly based IDS respond at inconsistent and irregular behavior of the system, on the basis of previous history. This IDS System matches the present profile with the previous profile of the system, when any significant divergence occurs, then the activity noticed like an attack. These are able to detect and recognize a Zero day attack. There are certain problems with current IDS that are huge amount of false positives, huge amount of false negatives, lack of efficiency, etc. To resolve these problems feature selection approaches are very essential in IDS systems.

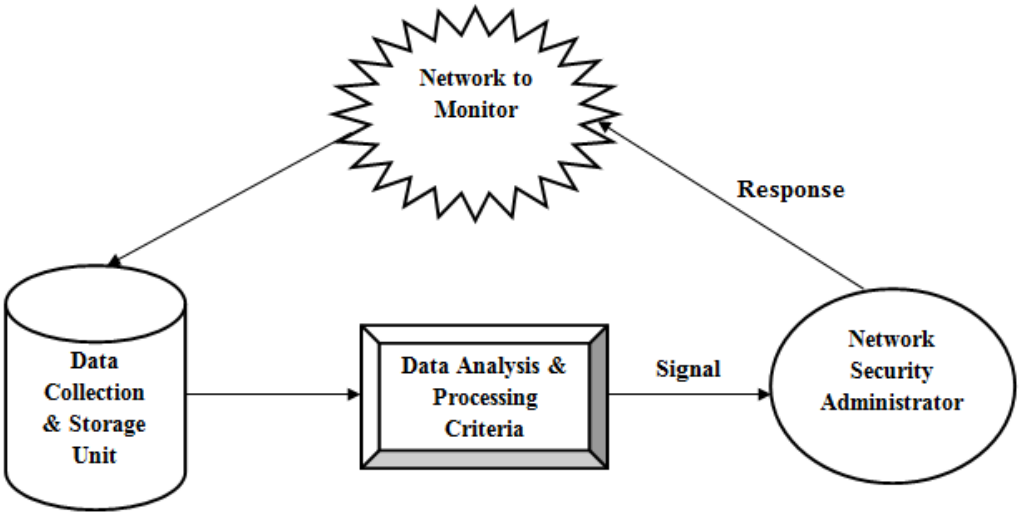

Fig 1: Structure of Intrusion Detection System

This paper is organized as follows: Initially in section 1 discussed the introduction part, section 2 discusses Feature selection approaches and classifiers, and section 3 describes Feature Selection - Related Work, in section 4 discussed the Research Methodology along with WEKA tool and NSL-KDD dataset. Results & Discussions with Performance Metrics have been discussed in section 5. Finally, section 6 concludes the research study.

2. FEATURE SELECTION

Feature selection performs an essential role in the intrusion detection process. Feature selection is a dynamic and fruitful area of research in the statistics, pattern recognition, data mining community and machine learning. The major purpose of feature selection is to select a best subset of input features and variables by eliminating the unusual features, which may be irrelevant or contain no more predictive information. FS reduces the amount of features that are irrelevant, noisy data or redundant data and bring the instant effects over the applications, such as speeding up data mining process, and enhance the mining performance like as result comprehensibility and predictive accuracy.

2.1 Feature Selection approaches

The Feature Selection approaches have main three forms that are Filter-based methods, Wrapper-based methods and Embedded methods.

2.1.1 Filter-based Method

The filter based methods used in feature selection approaches to find the best subset of features from the original data set. Filter based method uses the inherent characteristic of dataset to estimate the best feature from all set of data. To score the feature subset it uses the proxy measure rather than error rate. These methods are independent of the classification algorithm and provide better results in case of large data sets. These methods further classify as ranking and space search methods on the basis of strategy that they are following to select the best subset of features. In ranker based methods, every feature independently ranked by the uses of descriptive score functions and sorted in decreasing order on the basis of significance score. Ranker is much more efficient in computationally, but poor to examine redundant features. Space search methods do work with the idea of optimizing several distinct objective functions as use in the wrapper and embedded method.

2.1.2 Wrapper-based Method

The wrapper based method uses the predictive model for scoring the feature subset. All new subsets of features are used to train the model. It enhances the outcome of the particular predictors. It attains more identification rates rather than the filters. Wrapper has more computational cost in case of large data set and the possibility of the model to be over fitted.

2.1.3 Embedded Method

Embedded methods carry out feature selections through the learning of most favorable parameters. It is based on the performance estimation metric that is directly intended from the data set. Embedded method is a group of feature selection approaches among the learning process. It also depends upon the classification algorithm.

2.2 Need of Feature Selection Approaches

In IDS system has need of feature selection for simplification of the model, to reduce the training time and to improved the generalization. A feature selection approach produces the data mining, knowledge in more meaningful form. Feature Selection (FS) is required to reduce, all the dimensional space of the features. Increase the algorithm speed and overall reducing the complexity of the IDS system. FS is also needed to keep the original features as such and select from them a best subset of features, that are free from the noisy data, irreverent and redundant form of data. Further FS plays a great role in enhancing the quality of data and accuracy of the detection process. In experiment work, removed the feature set from the network traffic data set. Then applied the ranker based and best subset based feature selection approaches. To investigate intrusions 41 attributes contained data sets are considered. Main intention is to reduce the numbers of features that are useless and play no more important role in intrusion detection process. So, to execute the detection process in less time, with more accuracy feature selection approaches are more important.

2.3 Classifiers

Machine Learning algorithms are called Classifiers. Classifiers are used to classify network traffic dataset whenever apply over the dataset. They are applicable to

distinguish the network traffic into normal or abnormal (intrusive). In WEKA Environment tool has 76 classification algorithms, which are capable to perform task. We choose classifiers Bayes Net and Nave Bayes in our experiments.

2.3.1 Bayesian Network Classifier

Bayes Network Classifier used the various quality measures and searches algorithms with the base class. A Bayesian network (BN) (Pearl, 1988) consists of a directed acyclic graph G and a set P of probability distributions, wherever nodes and arcs in G symbolize the random variables and direct correlations between variables respectively, whereas a P is the set of local distributions for each node. Conditional probability table (CPT) normally specifies the local distribution in Bayes Net.

2.3.2 Naive Bayes Classifier

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.

3. FEATURE SELECTION - RELATED WORK

Mukherjee and Sharma (2012), [3] have proposed a new technique FVBRM model for reducing input features. Proposed technique further compared with CFS, InfoGain and Gain Ratio feature selection techniques. FVBRM method shows improvement in classification accuracy up to 97.78 %. Vitality of attributes is measured by the TPR classification accuracy and FPR of the network system. They compare only few numbers of feature selection techniques with proposed technique.

Singh et al. (2013), [4] focus on various existing feature selection techniques and evaluation is tested on the basis of three classifiers (Naive Bayes, J48 and PART) by using machine learning tools and Weka data mining over the UCI KDD CUP 1999 network traffic dataset. Filtered subset evaluator performs the best out of all techniques and reduces 82.93 % features with acceptable accuracy. Overall results indicate very less variations in accuracy by reducing the number of features.

Kumar et al. (2013), [5] propose the feature selection method of ranking and using the different feature selection algorithms like Info Gain, OneR, Gain Ratio, RELIEF etc. Combining the features of best algorithm those having high performance with use of J48 classifier on the KDDcup99 dataset. Experimental results indicate that the proposed model capable to reduces 70.73% in feature dimension space, near about 55-60% reduction in training time, and increased classification accuracy from 61.39% to 66.80% in detecting attacks. In this paper analyzes the results of only ranking based techniques.

Garg and Yogesh (2014), [6] have been performed the experiments to compare the various feature selection approaches. They compared the performance of the combination of six ranking based feature selection techniques by using Boolean AND operation over the Gain Ratio, Info gain, Chi square, Relief F, Symmetrical Uncertain and Filtered attribute evaluator techniques. The evaluation has been done using ten classification algorithms. Combination of Symmetric and Gain Ratio while considering top 15 attributes has maximum performance. At present work they have been focused only on Filter Model.

Kaur et al. (2015), [7] compared various feature selection techniques with the use of preprocessed NSL - KDD dataset. Various feature selection techniques like CFS using Best First & Genetic Search method, Filtered Attribute Method, Chi-square Attribute method, Info Gain method, Gain Ratio Method, Filtered Subset etc. KDD data set is used to reduce the training & test data sets. Naive Bayes Classifier is used to classify in this. Gain Ratio gives more accuracy (90.2567%) and Filtered subset (BFS method) takes less training time (0.21%). They perform experiments just at 37,789 records.

4. RESEARCH METHODOLOGY

4.1 WEKA

WEKA (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 modeling and data analysis. The functionality in WEKA is very easy due to it has a user graphical interface. Algorithms can be applied directly over the dataset or prepared by personal Java code. It provides 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 preprocessing tasks, classification, clustering, regression, visualization, and feature selection. It runs over any recent computing platform [8].

4.2 Network Traffic Dataset- NSL-KDD

Since 1999, KDD99 are most widely used dataset for the estimation of anomaly detection methods. The data set is organized by Stolfo et al. [9], and it is developed on basis of captured data in DARPA98 IDS evaluation program. The KDD training data set consists of about 4,900,000 single connection vectors all of which has 41 features and also labeled as an attack or normal, with precisely one particular type of attack. In these research experiments, selected 125973 connection records as a training data set and test data set. 60% data set used in training and remains 40% data set used in testing purpose. 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.

4.3 Experimental Setup

In this research, Methodology applied in experiments according to the work for comparing the existing different feature selection approaches. The main goal of this study is to analyze the effects of different features at the accuracy of the classifier to classify the various instances of the network traffic dataset, minimize the computational time and maximize the accuracy rate with lower number of features in threat detection. In the feature selection approaches firstly generate a subset of features by the use of forward addition and backward elimination process. Then evaluated the generated subset of features and compare it with the previous subset of the best features on the basis of some criteria. The process has completed only when the appropriate criterion and validated subset of features is achieved. In these experiments, nine feature selection approaches have been applied to decrease the dimensionality of data. Two classifiers are considered in analyzing the effects on accuracy. The

methodology is adopted as follows:

4.3.1 Selection approaches has done the Pre- processing of KDD dataset

4.3.2 To obtain the reduced feature set employing various feature selection approaches over the data set.

4.3.3 Split the training and testing dataset through a reduced feature set. Methodology is shown as in fig-2.

4.3.4 Creation of the trained model is done by using Bayes Net and Naive Bayes classifier with the help of reduced training data set.

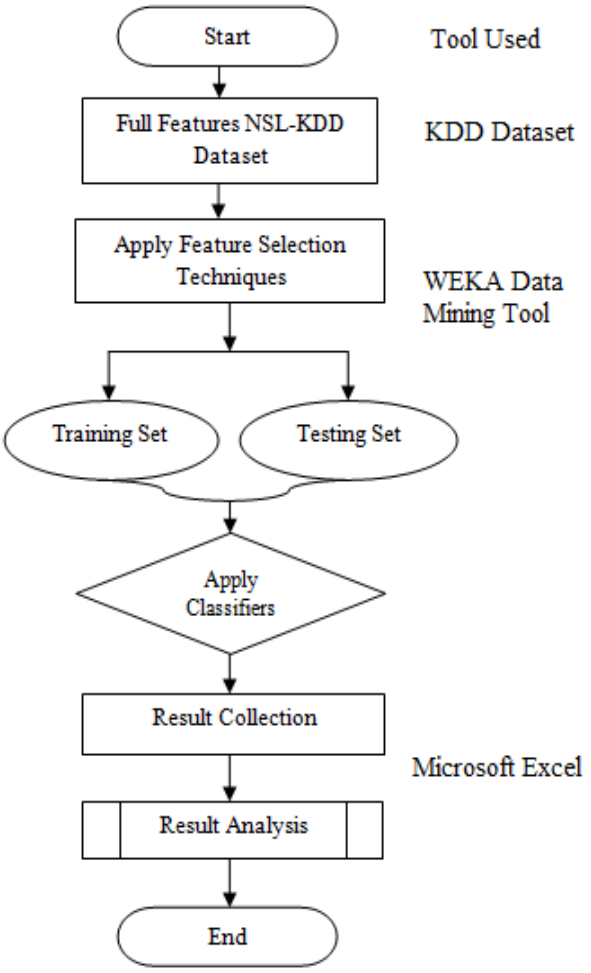

Fig 2: Research model used for experimentation

4.3.5 Computation of experimental results are depends upon the various performance metrics as like classification accuracy, TPR, FPR, Precision, Receiving Operating Characteristic (ROC) Area, F measure, Kappa Statistic and Training Time.

4.3.6 Comparative analysis has done at different feature selection approaches with search methods based upon different parameters for both Classifiers.

5. RESULT & DISCUSSIONS

In order to evaluate the existing feature selection approaches, have performed the experiments at the WEKA 3.6.9 toolkit on NSL-KDD dataset, were 125973 instances are used for training and testing purpose. The experiments carried out at the Intel Core i3 M 380 2.40 GHz processor with 2GB of RAM with 64 bit operating systems. In the feature selection approaches, to keep away the over fitting problem, apply 10 fold cross validation and it also improve the performance of the model and make it cost effective.

5.1 Performance Metrics

Different parameters used to analysis the results, which include classification accuracy, selected number of features, training time, True Positive (TP) rate and False Positive (FP) rate, Precision, F Measure, Kappa statics and ROC area. Table 1 show 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.

5.1.1 True negative

TN calculates the number of detected, normal instances which are normal in actuality.

5.1.2 False negative

FN calculates the number of detected, normal instances which are actually attacked.

5.1.3 True positive

TP calculates the number of detected attacks which are actually attacked.

5.1.4 False positive

FP calculates the number of detected attacks which are normal in actuality.

Table 1. Confusion Matrix

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

5.1.5 Classification 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.

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

5.1.6 True positive Rate (TP Rate)

TPR is defined as the ratio of number of classified attack connections and full amount of normal connections.

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

5.1.7 False Positive Rate (FP Rate)

FPR is defined as the ratio of the number of misclassified normal connections and full amount of normal connections.

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

5.1.8 Precision

This metric is defined with respect to the intrusion class. It should be high for more accuracy in IDS system.

$$Precision = \frac{TP}{TP+FP}$$

5.1.9 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.

5.1.10 F-Measure

The F-measure is defined as a weighted harmonic mean of recall and precision. It is high when both the recall and precision are high.

5.1.11 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.

5.1.12 Training time

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

5.2 Number of Selected Features

Table 2 shows the different features recommended by various feature selection approaches. In these experiments, have analyzed 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, Classification Accuracy, ROC Area, F Measure, Precision, Kappa Statistic 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, Consistency Subset Eval, Filtered Attribute Eval, Filtered Subset Eval, Gain Ratio, Info Gain, OneR with search methods Best First search, Greedy Stepwise, Genetic search, Linear forward search, Ranker and Rank search. These search methods seek for the set of all probable features in order to obtain a best subset of feature.

Table2. List of Selected Features by feature selection approaches

Feature Selection Technique	Search Method	No. of Attribute Selected	Selected Attributes
Full Features	Nil	41	All 41 Features
CFS Subset Evaluator	Best First Search	6	4,5,6,12,26,30
CFS Subset Evaluator	Genetic Search	15	4,5,6,8,10,12,17,23,26,29,30,32,37,38,39
CFS Subset Evaluator	Greedy Stepwise	6	4,5,6,12,26,30
CFS Subset Evaluator	Rank Search	12	3,4,5,6,12,25,26,29,30,37,38,39
Chi Squared Attribute Evaluator	Ranker	12	5,3,6,4,30,29,33,34,35,12,23,38,
Consistency Subset Evaluator	Greedy Stepwise	11	1,3,5,6,23,32,33,34,35,37,39
Consistency Subset Evaluator	Linear Forward Selection	10	1,3,5,6,23,32,34,35,37,39
Filtered Attribute Evaluator	Ranker	12	5,3,6,4,30,29,33,34,35,38,12,39
Filtered Subset Evaluator	Greedy Stepwise	6	4,5,6,12,26,30
Filtered Subset Evaluator	Best First Search	6	4,5,6,12,26,30
Gain Ratio Method	Ranker	12	12,26,4,25,39,30,38,6,5,29,3,37
Info Gain Attribute Method	Ranker	12	5,3,6,4,30,29,33,34,35,38,12,39
One R Attribute Method	Ranker	12	5,3,6,4,30,29,34,33,35,12,23,38
Symmetrical Uncer Method	Ranker	12	12,4,26,6,39,25,5,30,38,29,3,33

Table 3 Indicates the values obtained from different parameters through a high opinion of feature selection algorithm with a Bayes Net classifier. If carry the full feature dataset with full features (41), then the classifier gives an accuracy of 97.23% and time taken to build the model is 472.15 Sec. When the data set contains all the number of features, then computational complexity should be high. We have the obtained 98.49% accuracy by the use of the Consistency subset evaluator with Linear Forward Selection Search engine. It gives 0.998 ROC with a set of features 10, while the CFS Subset Evaluator with Best First search method takes a small amount of training time 38.89 Sec by using number of six features.

Table3. Comparative Analysis of different feature selection approaches with Bayes Net Classifier

Feature Selection Techniques	Search Method	No. of Selected Features	TP Rate	FP Rate	Precision	F Measure	ROC Area	Kappa Statistic	Accuracy	Training Time (Sec.)
Full Feature	Nil	41	0.972	0.031	0.973	0.972	0.998	0.9441	97.23%	472.15
CFS Subset	Best First Search	6	0.963	0.035	0.963	0.963	0.993	0.9255	96.29%	38.89
CFS Subset	Genetic Search	15	0.975	0.028	0.975	0.975	0.997	0.9493	97.48%	128.25
CFS Subset	Greedy Stepwise	6	0.963	0.035	0.963	0.963	0.993	0.9255	96.29%	43.86
CFS Subset	Rank Search	12	0.978	0.024	0.979	0.978	0.998	0.9563	97.83%	94.46
Chi Squared	Ranker	12	0.95	0.056	0.953	0.95	0.996	0.8994	95.02%	133.83
Consistency Subset	Greedy Stepwise	11	0.982	0.019	0.983	0.982	0.998	0.9648	98.25%	116.9
Consistency Subset	Linear Forward Selection	10	0.985	0.017	0.985	0.985	0.998	0.9697	**98.49%**	89.81
Filtered Attribute	Ranker	12	0.949	0.057	0.951	0.949	0.996	0.8967	94.89%	96.55
Filtered Subset	Greedy Stepwise	6	0.963	0.035	0.963	0.963	0.993	0.9255	96.29%	**41.73**
Filtered Subset	Best First Search	6	0.963	0.035	0.963	0.963	0.993	0.9255	96.29%	45.42
Gain Ratio	Ranker	12	0.978	0.024	0.979	0.978	0.998	0.9563	97.83%	97.13
Info Gain	Ranker	12	0.949	0.057	0.951	0.949	0.996	0.8967	94.89%	121.84
One R	Ranker	12	0.95	0.056	0.953	0.95	0.996	0.8994	95.02%	118.27
Symmetrical Uncer	Ranker	12	0.968	0.036	0.969	0.968	0.997	0.9349	96.77%	120.65

Table 4 indicates that the values obtained by the Naive Bayes classifier with respect to the different feature selection approach on the basis of various parameters. If carry the full dataset of features (41), then classifier provide accuracy of 90.25% and time taken to build the model is 485.81 Sec. While the Consistency subset evaluator with Greedy Search engine give 90.39% accuracy and 0.8054 ROC with a set of 11 features. The CFS Subset Evaluator with Greedy Stepwise search method takes a small amount of training time 53 Sec while using the number of features 6. Experimental results indicate that the consistency subset evaluator Method, Chi squared attribute evaluator Method, OneR Method outperforms from all the methods in conditions of classification accuracy.

Table4. Comparative Analysis of different feature selection approaches with Naïve Bayes Classifier

Feature Selection Techniques	Search Method	No. of Selected Features	TP Rate	FP Rate	Precision	F Measure	ROC Area	Kappa Statistic	Accuracy	Training Time (Sec.)
Full Feature	Nil	41	0.903	0.102	0.903	0.902	0.967	0.8034	90.25%	485.81
CFS Subset	Best First Search	6	0.82	0.204	0.852	0.813	0.937	0.6303	82.01%	56.84
CFS Subset	Genetic Search	15	0.898	0.112	0.905	0.898	0.955	0.794	89.85%	176.29
CFS Subset	Greedy Stepwise	6	0.82	0.204	0.852	0.813	0.937	0.6303	82.01%	**53**
CFS Subset	Rank Search	12	0.899	0.111	0.905	0.898	0.966	0.7951	89.90%	124.26

Chi Squared	Ranker	12	0.901	0.106	0.903	0.901	0.967	0.8002	90.12%	118.44
Consistency Subset	Greedy Stepwise	11	0.904	0.105	0.908	0.903	0.972	0.8054	**90.39%**	133.56
Consistency Subset	Linear Forward Selection	10	0.891	0.121	0.898	0.889	0.968	0.7777	89.05%	121.23
Filtered Attribute	Ranker	12	0.887	0.122	0.892	0.886	0.962	0.7714	88.72%	113.85
Filtered Subset	Greedy Stepwise	6	0.82	0.204	0.852	0.813	0.937	0.6303	82.01%	85.81
Filtered Subset	Best First Search	6	0.82	0.204	0.852	0.813	0.937	0.6303	82.01%	56.6
Gain Ratio	Ranker	12	0.897	0.114	0.903	0.896	0.965	0.7903	89.67%	117.1
Info Gain	Ranker	12	0.887	0.122	0.892	0.886	0.962	0.7714	88.72%	125.14
One R	Ranker	12	0.901	0.106	0.903	0.901	0.967	0.8002	90.12%	125.69
Symmetrical Uncer	Ranker	12	0.876	0.138	0.887	0.874	0.963	0.747	87.57%	126.89

6. CONCLUSION AND FUTURE SCOPE

Focus on research, compared the existing feature selection approaches. The main purpose of feature selection approaches is to reduce the useless, irrelevant, redundant and noisy feature from the network traffic dataset. The Feature Selection approaches are compared using various performance metrics like TPR, FPR, Classification accuracy, ROC Area, Precision, Kappa Statistic and Training Time. Full dataset contains all the features, it gives less accuracy and consumes more time to build the model. In this study, decided the best feature selection approaches on the basis of different performance metrics. CFS and Filtered subset evaluator take less training time with a small quantity of features by using Bayes Net Classifier. For obtaining the more accuracy with Bayes Net, we can use Consistency subset evaluator (Greedy stepwise search method). Naïve Bayes Classifier indicates that the Consistency attribute evaluator & Chi Squared attribute evaluator techniques provide more accuracy with the least number of features and CFS approach takes less training time with Rank search method. In future work will focus on enhancing the results of intrusion detection by combing the various feature selection approaches with a set of classiers.

7. REFERENCES

[1] D. Denning, An intrusion-detection model, Software Engineering, IEEE Transactions on SE-13 (2) (1987) 222-232.

[2] G. Stoneburner., Underlying technical models for information technology security., NIST Special Publication 800-33 , National Institute of Standards and Technology.

[3] S. Mukherjee, N. Sharma, Intrusion detection using naive bayes classifier with feature reduction, 2nd International Conference on Computer, Communication, Control and Information Technology(C3IT-2012) Procedia Technology 4 (2012) 119- 128

[4] R. Singh, H. Kumar, R. Singla, Analysis of feature selection techniques for network traffic dataset, in: Machine Intelligence and Research Advancement (ICMIRA), 2013 International Conference on, 2013, pp. 42-46.

[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] T. Garg, Y. Kumar, Combinational feature selection approach for network intrusion detection system, in: Parallel, Distributed and Grid Computing (PDGC), 2014 International Conference 2014, pp. 82-87.

[7] R. Kaur, G. Kumar, K. Kumar, A comparative study of feature selection techniques for intrusion detection, in:

[8] Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference 5 pp. 2120-2124.

[9] 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.

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