International Journal of Computer Networks and Communications Security

VOL. 4, NO. 1, JANUARY 2016, 1–11
Available online at: www.ijcncs.org
E-ISSN 2308-9830 (Online) / ISSN 2410-0595 (Print)

AdaBoost Ensemble Learning Technique for Optimal Feature Subset Selection

SAMAH OSAMA M. KAMEL[1], NADIA H. HEGAZI[2], HANY M. HARB[3], ADLY S. TAG EL DEIN[4], HALA HALA M.ABD EL KADER[5]

[1, 2] Electronic Research Institute, Department of Informatics

[3] Faculty of Engineering in Azahr University, Department of Computer Science

[4, 5] Faculty of Engineering in Benha, Department of Communication

E-mail: [1]samah_n2003@yahoo.com, [2]nhegazi@gmail.com, [3]harbhant@yahoo.com, [4]adlymerg@yahoo.com, [5]hala_mansour56@yahoo.co.uk

ABSTRACT

As a result of the spread of technology in the world, it must be secured to the communication data. In the recent year, the trend of various network organizations is the maintaining a high level of security to get the secure data communication. Data communication over the internet is exposed to a huge number of threats where the intrusion prevention system (IPS) can be used to detect many threats in real time. The classification of network events is an important part in IPS to detect threats. IPS performance is based on the dimensionality reduction of features. The feature selection is widely used in data mining approach. The proposed experiment identifies the best selected features by using Best First search technique in the wrapper model to obtain better performance of IPS. The proposed system shows the novel AdaBoost ensemble learner algorithm which consists of the number of the base learner algorithm. The base learner algorithm is learning Bayesian Network by using Genetic Algorithm global search approach to reduce the classification time of AdaBoost ensemble learner and improve the performance of BN. The proposed system consists of five stages; training dataset pre-processing, subset generation, model validation, model evaluation, and model comparison which will apply seven classification algorithms to evaluate on the training dataset to detect threats.

Keywords: *Intrusion Prevention System (IPS), Feature Selection, Wrapper Model, AdaBoost, Genetic Algorithm.*

1 INTRODUCTION

IPS is an advanced combination of IDS, personal firewalls and anti-viruses. It is not only to detect an attack, but also to stop it by responding automatically. The responsibilities of IPS are monitoring and analyzing events to identify undesirable activity, blocking malicious traffic, detecting worm and virus threats, changing the security environment. IPS Techniques are network based, wireless based, network behavior analysis and host based. There are three methodologies of IPS misuse detection, State-full protocol analysis and anomaly detection. Misuse or Signature based detection can detect many or all known attack patterns and tries to recognize known bad behavior. State-full protocol analysis is the process of comparing predetermined profiles of normal protocol activity for each protocol state against malicious observed events.

Anomaly based detection attempts to build models of "normal" behavior of a system by using data mining. It creates norms of activities which are used to detect anomalies that might indicate an intrusion. It tries to identify novel attacks by analyzing strange behavior from normal behaviors. The major problem in the real world classification technique is that gaining a better classifier algorithm with high performance. The classification of such intrusions plays a significant role to detect threats in the network intrusion prevention system. This process detects threats features which decrease

IPS performance. Due to the large size of data, it should remove unwanted features to reduce the number of irrelevant and redundant features, and decrease the running time of the learning algorithms to get better performance. The dimensionality reduction of features is widely used technique in data mining for reducing classification time and computational cost, increasing learning accuracy and building a better model. The dimensionality reduction is used to remove irrelevant and redundant features which can be divided into feature extraction and selection.

The feature selection selects a small subset of the relevant features of the original features which belong to different classes without any transforming and saves the physical meaning of the original features which leads to higher accuracy rate (AR), lower computational cost and better model interpretability. The process of feature selection begins with the subset generation; a feature subset will be selected by using a certain strategy and send to the second step which called subset evaluation. Subset evaluation; it is evaluated according to certain evaluation criterion. After the evaluation process, the best evaluated subset will be selected from all subsets which called stopping criterion. If the result is satisfied, it will be stopped. Otherwise, it will be continued. The selected subset will be validated by using a validation set and this step is called result validation. There are different algorithms of feature selection which can be categories into supervised, unsupervised and semi-supervised algorithms. Supervised feature selection algorithms can be divided into three models that are filter model, wrapper model and embedded model. In the filter model, the feature selection is separated from classifier learning algorithms which the learning algorithms bias doesn't interact with a feature selection algorithm bias. It depends on the measures of the characteristics of the training data such as distance, consistency, dependency, information, and correlation. The disadvantage of the filter model is that it completely ignores the effects of the selected feature subset on the performance of the predetermined learning algorithm. The wrapper model uses the performance of the learning algorithms to evaluate the quality of the selected features. It is a suitable method to define the feature selection problem. The wrapper model contains three components that are feature selection search, feature evaluation and induction algorithms. The feature selection search produces a feature set and the feature evaluation uses the learning classifier to evaluate its performance. This estimation will be returned back the feature selection stage to implement the next iteration of feature subset selection. The final feature set is selected based on the highest estimated value of the feature set and then it learns the classifier algorithm.

The result of the learning classifier algorithm is estimated on the testing set. The disadvantage of wrapper model is that it must run the classifier many times to assess the quality of selected subsets of features so it is very computationally cost but it gives high performance. There are many search techniques in the wrapper model such as hill-climbing, genetic algorithm and best first. The hill-climbing technique expands the current feature of subset and transports it to the subset which is the highest accuracy but without improving over the subset feature. A genetic algorithm is a computer simulation which represents chromosome of candidate solution to be the optimal solution. Each solution set is called population, which consists of vectors (individual). Each item in vectors called genes. An individual represents features which encoded 0 (none selected) or 1 (selected). The disadvantage of genetic algorithm technique is that the genetic algorithm produces a subset of the selected feature but it doesn't produce better performance. Best first search selects the desired set that has not been expanded which leads to be a more robust method than hill- climbing. The wrapper model produces better performance than filter model, but it produces high computational cost. Wrapper model obtains maximum quality of the feature selection. Embedded Model combines the advantages of filter model and wrapper model. It uses the performance of earning algorithms to evaluate the selected features, quality and less computationally expensive.

This paper will display the proposed system which will apply classification algorithms to KDD and NSL-KDD dataset. The experiment goal is that classify and identify unknown record in the test dataset to deal with an attack and prevent it. After we removed the redundant and duplicated records, we get the different types of threats. The training dataset will give a broad diversity of intrusions and normal activities which is simulated in the network environment and indicates the legitimate network traffic. The proposed system will use different classifier algorithms to select the best features and evaluate the performance metrics of each classifier algorithm.

2 RELATED WORK

[1] This research used NSL-KDD data set and implemented the evaluation of seven trees based classifier algorithms to classify network events

based on supervised discretization and feature selection. The experiment is based on error metrics. Feature selection is carried out by using consistency subset (CONS) evaluator and correlation feature subset (CFS) evaluator which selected seven features.

[2] The author used CFS Subset to reduce the dimensionality of the data set in 15 features by using different classification algorithms which are Random Forest, J48, SVM, CART and Naive Bayes for the NSL-KDD data set with and without feature reduction. The result of the experiment has been displayed that the accuracy rate of using 15 features is higher than the accuracy rate of 41 features.

[3] The authors provide a relevant set of features for detection of DOS attacks and analyzed NSL KDD dataset with 6 fold' cross validation. The experiments have been implemented by using 3 features set which included 41 features, 28 features and 8 features by using C4.5 decision tree classification algorithms. The authors measured the classification time and accuracy rate of the three features set. The experiment has been displayed that the accuracy rate of 8 features set is greater than the two feature sets. The classification time of 8 features set is less than the two feature sets.

[4] The authors have been implemented the proposed feature selection methods using AR and compared it with three feature selectors CFS, IG, and GR. The experiment displayed the detection rate of our method is higher than the detection rate of full data and is also as high as the detection rate of other methods. The false alarm rate is lower than full data and is as low as false alarm rate of other methods.

[5] The authors evaluated the performance of the different of learning algorithms which has been used to rank the 41 features of NSL-KDD data set. The feature selection method is wrapper model. The experimental results of the selected subset of features reduced the input data with a 80 % reduction in time. The experiment has been displayed that the intrusion detection system with a feature selection algorithm has better performance that that without the feature selection algorithm.

3 THE PROPOSED SYSTEM

Our experiment is based on a wrapper model which uses the performance of the learning algorithms to evaluate the quality of the selected features. The wrapper model consists of three components that are feature selection search, feature evaluation and induction algorithms. The best first search technique is used in our experiment because it is more a robust search technique for feature selection process.

As shown in Figure 1, the feature selection process is implemented in five steps. The first step is applying Supervised attribute filtering discretize technique for data pre-processing to obtain more efficient data set and improve the performance of IPS. In supervised attribute filtering discretize, it transforms a range of numeric attributes in the dataset into nominal attributes and it is very flexible.

The second step is subset generation which contains three stages that are implementing learning classifier algorithms, applying best first search technique and selecting the highest three features. The first stage is implementing learning classifier algorithms by using seven algorithms These classifiers are chosen according to the performance parameters that are used to evaluate each algorithm and produce a good performance. The performance metrics include accuracy rate (AR), true positive rate (TPR), false positive rate (FPR), precision rate, mean absolute error (MAE), Kappa Statistics (KS), confusion matrix and classification time. AR is the performance of the system which is evaluated by calculating the correctly classified records ratio to the total of records. TPR is also sensitivity, hit-rate or Recall which is evaluated by calculating true positive to the total number of true positive and false negative. FPR is the ratio of incorrectly classified normal record of the total of true negative and false positive. Precision rate is the ratio of true positives to the combined true and false positive. MAE specifies how the predicted values are different from the actual values. KS measures the consistency between the predicted values and the actual values in a dataset. If the higher of KS is obtained, the higher of consistency and the algorithm performance is obtained. The second stage is applying best first search technique. The seven classifiers are used to implement feature selection search by using the best first search technique which is a robust method. The best three features are selected which is called selecting the best three features stage.

The third step is subset evaluation, which contains two stages. The first one is applying the classifiers to generate the subset and the second is generating the model validation by training dataset. In this step; we will apply the novel AdaBoost ensemble learning algorithm which produces minimum classification time and high performance.

The fourth step is to model evaluation by calculating the performance metrics. The final step in the proposed system is stopping criterion. After the subset evaluation process, the best evaluated

subset will be selected from all subsets. If the best performance of the model is satisfied, it will be stopped. Otherwise, it will increase the number of features by one feature. By implementing the performance model comparison, the optimal feature subset is detected which called result validation.

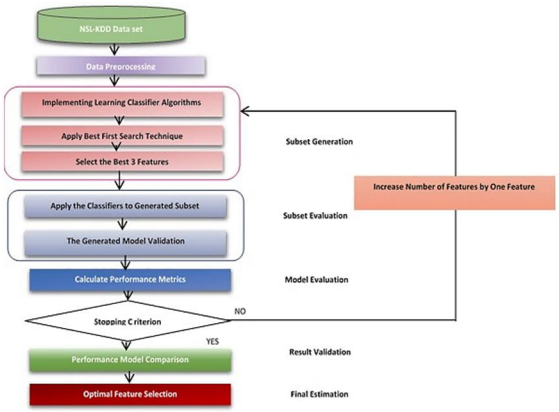

Fig. 1. The Proposed System

4 CLASSIFIER ALGORITHMS

Seven classifier algorithms are used to implement our experiment. These algorithms are used to evaluate each algorithm and produce a good performance. The classification algorithms are Stacking, Bagging, Random Forest Tree, PART, IBK, Bayes Network and Novel AdaBoost Ensemble Learning Algorithm.

4.1 Random Forest Tree (RFT)

Random Forest Tree (RFT) is an ensemble classifier which consists of individual decision trees. RFT combines bagging idea and a random selection of features which is independent of the structure a collection of decision trees with controlled variation. RFT is one of the highest accurate classifier among classification algorithm.

4.2 PART

PART algorithm is one of the decision rules and it has the highest performance among the decision rules. It uses separate and conquers. In each iteration, it obtains the best leaf when it builds a fractional of C4.5 decision tree.

4.3 Stacking

Stacking learns classifier in sequence process and it used to improve the performance of the ensemble by correcting the errors. It handles the bias classifier and using these biases to improve classification. Stacking uses a base classifier to correct the errors of a previous classifier. It learns classifier in sequence process. It is implemented in two levels. First level is that it generates a number of first-level individual learners from the training data set by using different learning algorithms. Second is that those individual learners are then combined by a second level learner, which is called as a meta-learner.

4.4 IBK

IBK is a lazy classifier algorithm which makes the use of the k-nearest-neighbor classifier. k-NN is a type of instance-based learning, or lazy learning. The principle of k-NN is that the input space is similar to the output space and stores the training set. When it trains the test set, k-NN identifies k instances from training set which is close to the majority class.

4.5 Bagging

Bagging technique learns the classifier in a parallel process. The benefit of parallel ensemble methods is gaining classification speed which can be used in multi core processors. It is called Bootstrap sample which creates novel training sets by random sampling. It consists of an independent learning algorithm to reduce error. Bagging trains each model in the ensemble by using a random subset of the training set. The train dataset contains m number of training records. Sample of m training records will be generated by sampling with replacement. Some of original training records appear more than once while others don't appear in the sample. The process applies t times to obtain samples of m training records. After that, a base learner can be trained by applying learning algorithm in each sample. The final output of the learning algorithm is the prediction result which is combined by majority voting. The Bagging technique advantages are that it is fast, simple and easy implementation. Bagging reduces over-fitting of the training data.

4.6 Bayes Network

A Bayesian network (BN) BN represents the joint probability distribution in discrete, continuous and hybrid environment. BN consists of a number of nodes and edges. Nodes represent random variable. The edges represent statistical dependencies. BN structure relies on the dependence structure among large no. of variables. Dependence structure implements the probability conclusion of these variables in an efficient manner. BN is used to calculate the conditional probability of one node, which gives certain values of the other nodes. BN consists of two components; a directed acyclic graph (DAG) representing the dependency structure among the variables in the network and a set of conditional probability table (CPT) for discrete data or a probability density function for continuous data. BN learning can be categorized into two approaches that are parametric and structural learning. Parametric learning is used for learning the parameters when the structure is known. Structural learning identifies the topology of the BN which gives the best description the observed data. Structural learning is implemented by two stage process; learn a network structure, then learn the probability tables. The structural learning can be divided into score and search-based approach which learns the network structures and constraint-based (CB) approach which learns the edges composing a structure. There many methods of Score and search based methods such as score and search methods. The proposed system will focus the search method because it is more intelligent, which move through the space of possible networks. It discovers the best network from the set of all networks which each node has no more than a number of parents. The implementation of search method takes more time. Search approach can be categorized into two approaches that are local search and global search. The difference between the local and global search that the global search uses cross validation. Cross validation process splitting the data set into ten parts and takes one part as a validation set and trains remain parts. This process is repeated in another part which provides high quality of BN.

4.7 Novel AdaBoost Ensemble Learning Algorithm

Genetic algorithm is categorized as a global search heuristic approach which is used to search for an optimal subset of predictor variables for improving BN classifier. Genetic search has been implemented through a population of BN structure. The BN structure represents an array of the node number. A population of the BN structure represents a set of solution which is represented in binary strings. A population of the BN structure consists of a number of vectors, which called chromosome or individual which represents the features. The process of the genetic search process is implemented through four the steps which are initialization, selection, crossover (recombination) and mutation. The figure 2 represents the BN learning algorithm uses genetic search.

1. Initialization: An initial population of BN structure is generated randomly from many individual solutions which covers the entire range of the search space (solutions).

2. Selection: For each successful generation, a rate of the existing population is selected to produce a novel generation. The best individual solutions are selected by evaluating the fitness which should be fitted with a population of BN structure.

3. Crossover: In each generation, the modified individuals reproduce by applying crossover and mutation to create one or more offspring which is added to the population. Cross over analysis the distinct the bit representation of BN structure and combine them to form a novel population (solution). The combination is implemented by that children take a random first bit from parent and adding the remainder of the other parent to create offspring.

4. Mutation: After crossover, we have a novel population with full of individuals. The implementation of mutation is used for avoiding the repeated individuals. Mutation flips a bit in the bit representation of the BN structure to change it by a small amount or replace it with a novel value. Genetic search approach terminates after a specific number of iterations, when we get the maximum fitness

level of BN structure populations which represent the satisfaction solution. The final algorithm is Bayesian Network learning by using GA search method.

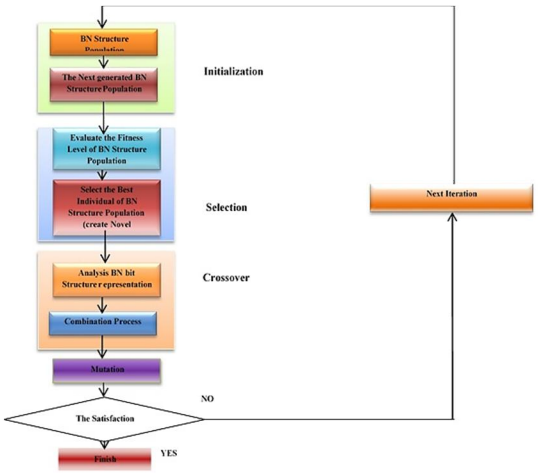

Fig. 2. The BN learning algorithm uses genetic search.

The BN learning based on GA search method is used as a base learner algorithm to implement AdaBoost techniques to produce a novel AdaBoost ensemble learner algorithm. The novel AdaBoost ensemble learner consists of ten base learners. The novel AdaBoost ensemble learner trains sequentially the classifiers. The process has been implemented in 10 rounds. For each round has been performed as the following steps:

- The classifier analyses all the records in the dataset and assigns the equal weights to all training examples. The usage of the weights is that allow the algorithm to classify records which were difficult to classify.
- Then classifier trains the records of the training dataset. The records that have higher weight are the ones which were classified wrongly by algorithm. If the error rate is over 50 %, it will be repeated this step. Otherwise, it will be moved to the next step.
- The classifier determines the weights and then updates all weights to reduce the processing time.
- The predictions of the algorithms are combined into a single prediction which represents the classification accuracy.

Input: Data set D = {(x1, y1), ……, (xn, yn)}
Number of iterations T
An initial distribution W_0 (i) the instance.
For t=1 to T

Train a base learner:

$$ht(x) = argmin \sum_{i=0}^{n} w(t-1)^i \quad (1)$$

Calculate the error rate of hi:

$$\square t = \frac{\sum_{yi \neq hi(xi)} w(t)^i}{\sum_{i=1}^{n} w(t)^i} \quad (2)$$

Calculate the Coefficient vector of base learner:

$$\alpha t = \frac{1}{2} \ln[\frac{(1-\square t)}{\square t} \quad (3)$$

Calculate the normalization factor:

$$Z_{t+1} = 2\,(1 - \square t) \quad (4)$$

Calculate the Weight of each record:

$$W_{t+1}^{i} = \frac{[W_t^i \exp[\,\alpha t\; yi\; ht(x)]}{Zt} \quad (5)$$

If error rate

$$\begin{cases} > 50\,\%\,, it\ will\ be\ repeated\ this\ step \\ < 50\%,\ it\ will\ be\ moved\ to\ the\ next\ step \end{cases}$$

End

Output:

$$H(x) = \sum_{k=1}^{T} \alpha_k\, h_k(x) \quad (6)$$

Where: X_(n): an object or instance, y_i: the classification output (association class), T: the number of iterations (t = 1,….,T), K: the number of base learner, n: the number of instance, W: the Weight of each record, □: the error rate, A: Coefficient vector of base learner, h(xi): the hypothesis of a base learner, H(x): the final combination which represents the set of hypothesis, Z_t: the normalization factor which enables weight distribution to be distributed.

5 EXPERIMENTAL SETUP

Our experiments have been used to run with the used platform which is Intel Core i7 4500CPU, 2.40GHz and RAM is 8GB and MS windows 8 professional 64 bits. The development environment is Waikato Environment for Knowledge Analysis (Weka) version 3.7.12 which is an open source machine learning package. Weka applications are the Explorer, Experimenter, Knowledge Flow and Simple CLI. For Explorer section, it contains tools for data pre-processing, classification, clustering, association rules, select attributes and visualization. In this search, our experiments have been implemented in five steps that are the training dataset pre-processing, subset generation, model validation, model evaluation and model comparison.

5.1 Training Dataset Pre-processing

The proposed system has been performed by using KDD and NSL-KDD dataset. The experiment goal is that classify and identify unknown record in the test dataset to deal with an attack and prevent it. After we removed the redundant and duplicated records, we get the different types of threats. The

training dataset gave a broad diversity of intrusions and normal activities which is simulated in the network environment and indicates the legitimate network traffic. The training and test dataset belong to one of the following five categories: Normal, DoS (denial of service), R2L (root to local), U2R (user to root) and Probing (surveillance). Every attack category gives different types of attacks. The numbers of instances in the training data set are 48916 instances. The training dataset includes 41 features which can be categorized into three types; numeric (or continuous), nominal (or symbolic), and binary (or discrete). Supervised attribute filtering discretize technique is applied to the training dataset to obtain more efficient data set which leads to improve the performance of learning algorithms. Supervised attributes filtering discretize are simple, flexible and gives better classification results.

5.2 Subset Generation

In this section, the first step of feature selection is applied by using seven classifier algorithms. The wrapper model is applied on the 41 features of the raining data set. The best first search technique is implemented to obtain the best first features in the training dataset. The simulation results are shown in the following table.

Table 1: Best First feature selection

Classifier Algorithm	Feature Number
Novel AdaBoost Ensemble Learner	2,3,4,5,7,12,23,25,29,30,32,33,35,36,37,40,34,1
Stacking	5,7,29,35,36,3,32,37,12,30,33,35,2,3,4,23,37,40,1,34
Bagging	5,7,12,2,4,3,30,37,36,32,40,29,35,33,23,25,29,40,18
RFT	5,7,12,2,4,3,30,37,36,32,40,29,35,23,25,29,40,33
PART	5,7,12,2,4,3,36,37,23,32,30,35,29,25,23,40,33,35
Bayes Network	2,3,4,5,30,32,35,37,7,1,6
IBK	5,7,12,2,4,25,30,37,40,3,32,33,29,35,36,23

We concluded from the previous table that best first feature set is 2, 3, 4, 5, 7, 12, 23, 25, 29, 30, 32, 33, 35, 36, 37 and 40. The best first selected features are protocol_type,service, flag,src_bytes, land, logged_in, count, serror_rate, same_srv_rate,diff_srv_rate,dst_host_count,dst_host_srv_count,dst_host_diff_srv_rate,dst_host_same_src_port_rate, dst_host_srv_diff_host_rate and dst_host_rerror_rate which obtained by the classifier algorithms. We will select the best first three features as the generated subset and apply the model validation and model evaluation processes. If the result is satisfied, it will stop. Otherwise, it will be increased number of features by one feature.

5.3 Model Validation

The model validation process is the next step of the experiment. The classifier algorithms will be applied to the first generated subset. We use the state of the art train set process which train the training data set by using seven learning classifier algorithms. Then, the model evaluation process will be applied in the next section.

5.4 Model Evaluation

The performance metrics are an indicator of this experiment to evaluate the performance of classifier algorithms. The classification accuracy is based on the confusion matrix which explains that how many instances of each class are correctly and incorrectly classified. The incorrectly instances lead to high false alarm rate. If the confusion matrix will increase, the accuracy rate will decrease. So it should be as minimum as possible to obtain high accuracy rate. The classification time plays an important role in the performance of the algorithm. It should be low to obtain high performance.

5.5 Model Comparison

The final step in the proposed system is the model comparison which compares the performance metrics of seven classifier algorithms. When the AR, TPR, precision rate and KS increase, the performance increase. When FPR MAE, classification time and confusion matrix decrease, the performance decrease and the learning classifier algorithms will be more efficient. All these comparison parameters are the most major criteria to indicate the performance of learning classifier algorithms. The best performance of algorithm indicates that the selection of feature subset is the best.

6 EXPERIMENTAL RESULTS AND DISCUSSION

The table 2 and figure 3 show the evaluation of accuracy rate for seven learning classifier algorithms which apply to the best first features. This table represents the relation of AR and number of features. If the result is satisfied, it will stop. AR is increased by increasing number of features. The feature number from 9 to 13 produced the same results. From the table 2, we deduced the following result: -

- Novel AdaBoost and BN, AR of 16 features is greater by 0.1 % and 0.2 % than the accuracy rate of 41 features respectively.
- RFT and IBK, bagging, AR of 16 features is equal to the accuracy rate of 41 features which is equal 99.9 %.
- PART, Stacking, AR of 16 features is less by 0.1 % and 0.04 %, than the accuracy rate of 41 features respectively.

Table 2: AR of best first 16 features

No.	Novel AdaBoost	PA.	RFT	BN	Sta.	IBK	Bag.
3	94.59	94.52	94.5	94.4	94.5	94.5	94.5
4	98.62	99.03	99.1	98.5	99	99.1	99.1
5	98.65	99.0	99.2	98.5	99	99.2	99.1
6	98.91	99.0	99.2	98	99.1	99.2	99.1
7	99.46	99.4	99.67	98.5	99.2	99.3	99.6
8	99.46	99.5	99.71	98.3	99.4	99.4	99.6
9	99.4	99.5	99.7	98.54	99.42	99.72	99.6
14	99.8	99.7	99.9	98.7	99.6	99.9	99.8
15	99.8	99.9	99.9	98.9	99.9	99.9	99.9
16	99.9	99.8	99.9	98.9	99.9	99.9	99.9
41	99.8	99.9	99.9	98.7	99.94	99.9	99.9

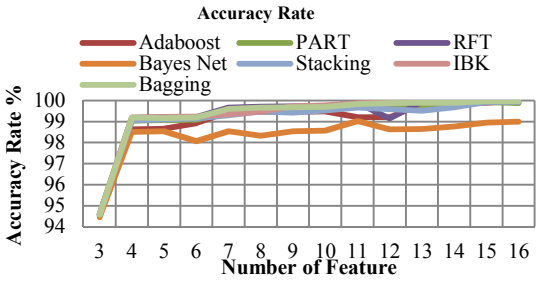

Fig. 3. AR of best first 16 features

The table 3 shows the evaluation of KS for seven learning classifier algorithms. We selected the best first three features and are used as a generated subset. The features added till the best result is detected. When we increased the number of features, KS was gradually increased. The figure 4 illustrates the gradually increasing in KS. From the table 3 we deduced the following result:

- Novel AdaBoost and BN, KS of 16 features is greater by 0.28 % and 0.4 % than KS of 41 features respectively.
- RFT and IBK, the KS of 16 features is equal to the accuracy rate of 41 features which is 100%.

• PART, Stacking and Bagging, KS of 16 features is less by 0.06 %, 0.02 % and 0.01 % than 41 features respectively.

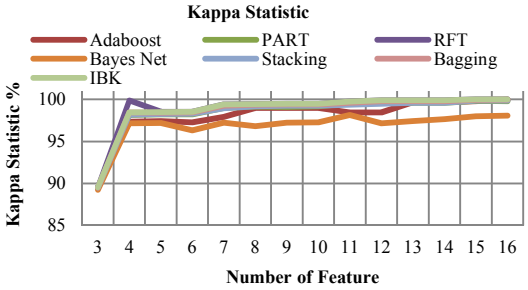

Fig. 4. KS of best first 16 features

The table 4 shows the evaluation of MAE for seven learning. MAE is gradually decreased with increasing number of features. The figure 5 illustrates the gradually decreasing MAE. From the table 4 we deduced the following result: -

- Novel AdaBoost and BN, MAE of 16 features is less by 0.02 % and 0.01 than features respectively.
- RFT, IBK and Bagging, MAE of 16 features is equal to MAE of 41 features.
- PART and Stacking, MAE of 16 features is greater by 0.01 % than MAE of 41 features.

Table 3: KS of best first 16 features

No.	Novel AdaBoost	PA.	RFT	BN	Sta.	IBK	Bag.
3	89.5	89.3	89.5	89.2	89.4	89.53	89.51
4	97.3	98.1	99.8	97.15	98.1	98.44	98.44
5	97.4	98.2	98.5	97.2	98.2	98.5	98.4
6	97.2	98.2	98.5	96.3	98.3	98.53	98.44
7	97.9	98.9	99.3	97.2	98.9	99.39	99.25
8	98.9	99.1	99.4	96.8	99.1	99.46	99.37
9	98.9	99.09	99.4	97.2	99.2	99.48	99.39
14	99.6	99.6	99.8	97.6	99.5	99.9	99.79
15	99.8	99.84	100	98.01	99.8	100	99.95
16	99.97	99.79	100	98.09	99.8	100	99.95
41	99.6	99.85	100	97.69	99.8	100	99.96

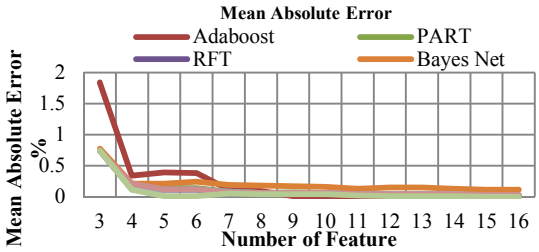

Fig. 5. MSE of best first 16 features

Table 4: MAE of best first 16 features

No.	Novel AdaBoost	PA.	RFT	BN	Sta.	IBK	Bag.
3	1.84	0.76	0.74	0.77	0.74	0.74	0.74
4	0.34	0.14	0.12	0.21	0.13	0.11	0.21
5	0.39	0.14	0.11	0.21	0.12	0.01	0.12
6	0.38	0.14	0.11	0.24	0.12	0.01	0.12
7	0.12	0.08	0.06	0.19	0.08	0.05	0.07
8	0.09	0.06	0.05	0.18	0.06	0.04	0.06
9	0.01	0.07	0.05	0.17	0.06	0.04	0.06
14	0.03	0.03	0.03	0.13	0.04	0.01	0.04
15	0.02	0.01	0.02	0.11	0.01	0	0.03
16	0	0.02	0.02	0.11	0.01	0	0.03
41	0.02	0.01	0.02	0.12	0.01	0	0.03

The table 5 and figure 6 show that the precision gradually increases with increasing number of features. From the table 5 we deduced the following result:

- Novel AdaBoost and BN, the precision of 16 features is greater by 0.1 % than the precision 41 features.
- PART, RFT, Stacking, IBK and Bagging, the precision of 61 features is equal to the precision of 41 features.

Table 5: Precision of best first 16 features

No.	Novel AdaBoost	PA.	RFT	BN	Sta.	IBK	Bag.
3	94.8	94.7	94.8	94.6	94.6	94.8	94.8
4	98.6	99	99.2	98.5	99	99.2	99.2
5	98.7	99.1	99.3	98.5	99.1	99.3	99.2
6	98.6	99.1	99.3	98.2	99.1	99.3	99.2
7	99.2	99.5	99.7	98.6	99.3	99.4	99.6
8	99.5	99.6	99.8	98.5	99.5	99.8	99.7
9	99.5	99.5	99.8	98.7	99.4	99.8	99.7
15	99.9	99.9	100	99.3	99.9	99.99	100
16	100	99.9	100	99.3	99.9	100	100
41	99.9	99.9	100	99.2	99.9	100	100

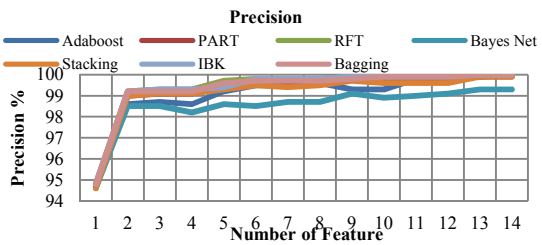

Fig.5. Precision of best first 16 features

The table 6 and 7 show the evaluation of FPR and TPR for seven learning classifier algorithms If the number of features increases, FPR decreases and TPR increases. The figure 7 and 8 illustrate the gradually decreasing in FPR and increasing TPR. We noted that all FPR and TPR of 16 features are equal to 41 features.

Table 6: FPR of best first 16 features

No.	Novel AdaBoost	PA.	RFT	BN	Sta.	IBK	Bag.
3	1.7	1.7	1.7	1.7	1.7	1.7	1.7
4	0.6	0.6	0.5	0.07	0.5	0.5	0.05
5	0.5	0..5	0.4	0.06	0.4	0.4	0.05
6	0.1	0.5	0.4	0.06	0.5	0.4	0.04
7	0	0.1	0.1	0.02	0.2	0.2	0.01
8	0	0	0	0.01	0.1	0	0
9	0	0	0	0.01	0.1	0	0
15	0	0	0	0.01	0	0	0
16	0	0	0	0.01	0	0	0
41	0	0	0	0.01	0	0	0

Table 7: TPR of best first 16 features

No	Novel AdaBoost	PA.	RFT	BN	Sta.	IBK	Bag.
3	94.6	94.5	94.6	99.4	94.6	94.6	94.6
4	98.6	99	99.2	98.5	99.1	99.2	99.2
5	98.7	99.1	99.2	98.5	99.1	99.2	99.2
6	98.6	99.1	99.2	98.1	99.1	99.2	99.2
7	98.9	99.5	99.7	98.5	99.3	99.3	99.6
8	99.5	99.6	99.7	98.3	99.5	99.7	99.7
9	99.5	99.5	99.7	98.5	99.4	99.7	99.7
14	99.8	99.8	99.9	98.8	99.7	99.9	99.9
15	99.9	99.9	100	99	99.9	99.99	100
16	100	99.9	100	99	99.9	100	100
41	99.8	99.9	100	99.1	99.9	100	100

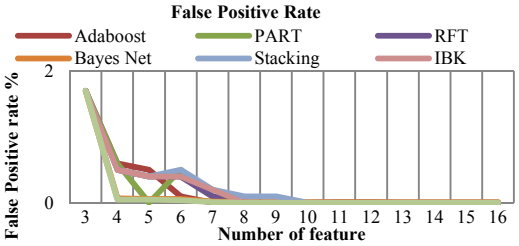

Fig.7. FPR of best first 16 features

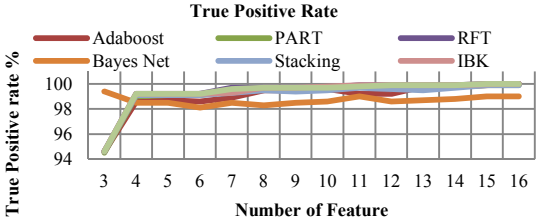

Fig.8. TPR of best first 16 features

Table 8 the classification time will increase by the increasing number of the feature. The figure 9 illustrates the gradually increasing in the classification time.

Table 8: The Classification Time of best first 16 features

No	Novel AdaBoost	PA.	RFT	BN	Sta.	IBK	Bag.
3	39.46	1.43	9.35	0.24	304.37	1200	82.68
4	40.49	4.2	10.14	0.27	364	1220	99.06
5	48.47	3.1	13.05	30	370.85	1260	178.11
6	58.84	3.24	13.9	0.32	372.22	1360	179.22
7	60.3	5.3	14.22	0.58	380.56	1410	196.75
8	61.01	5.8	14.33	0.91	394.56	1440	197.3
9	62.39	7.07	21.87	0.99	394.84	1460	198.4
14	73.48	8.9	28.54	2.43	400.57	1623	204.8
15	76.32	10.46	30.26	2.48	410.5	1652	206.8
16	194.02	12.86	32.47	2.64	420.61	1674	210.69
41	959.66	37.12	54.93	3.19	573.85	2913	425.46

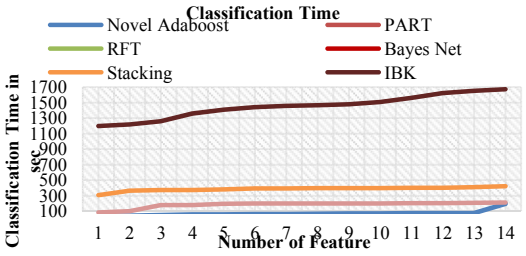

Fig.9. The Classification Time of Best First 16 Features

The incorrectly instances lead to high false alarm rate which gives high confusion matrix. It should be as much less as possible to obtain high performance of the algorithm. number of features is inverse proportion with confusion matrix as shown in figure 10 and 11.

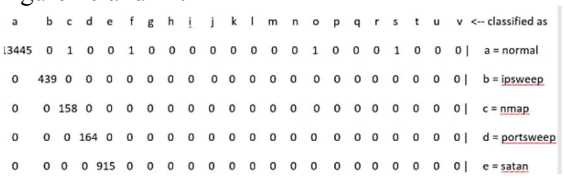

Fig.10. The Confusion matrix of best first 3 features

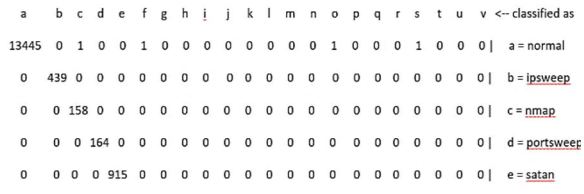

Fig. 11. The Confusion Matrix of the best first 16 features

7 CONCLUSIONS

Our experiments have been implemented in five steps that are training dataset pre-processing, subset generation, model validation, model evaluation and model comparison. This experiment has been implemented by using best first search technique in a wrapper model to select the significant features by using seven classifier algorithms. These classifier algorithms produced best 16 features set. The proposed system has been presented novel AdaBoost ensemble learning algorithm which is based on BN learning using GA search method. The novel AdaBoost ensemble learner algorithm improved the performance metrics of BN learning using GA search method which introduced high accuracy rate (AR), precision rate, recall, KS, and confusion matrix and MAE. The novel AdaBoost ensemble learner algorithm reduced the classification time, which can be used for an adaptive technique where the other AdaBoost technique consumes much more time for learning dataset. The reduction of classification time increased the training speed so it may be used in the multi-core architecture. When we used GA for learning the BN, it will prevent the overfitting of training data. The structure of novel AdaBoost ensemble learner algorithm represented the correlation among the attributes of the dataset.

8 REFERENCES

[1] Sumaiya Thaseen and Ch. Aswani Kumar, "An Analysis of Supervised Tree Based Classifiers for Intrusion Detection System", Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), Salem, 21-22 Feb. 2013, pp. 294-299.

[2] S. Revathi and A. Malathi, "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 12, December – 2013, pp.1848-1853.

[3] Pooja Bhoria and Kanwal Garg, "Determining feature set of DOS attacks", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, Issue 5, May 2013pp. 875-878.

[4] Sang-Hyun Choi and Hee-Su Chae, "Feature Selection using Attribute Ratio in NSL-KDD data", International Conference Data mining, Civil and Mechanical Engineering

(ICDMSME'2014), Bali (Indonesia), Fecb 4-5, 2014, pp.90-92.

[5] Hany M. Harb, Afaf A. Zaghrot, M. A. Gomaa and Abeer S. Desuky, "Selecting optimal subset of features for intrusion detection systems", Advanced in computational sciences and technology, Vol. 4, 2011, pp. 179-192.

[6] Data Classification: Algorithms and Applications, Charu C. Aggarwal, CRC press, chapter, Jiliang Tang, Salem Alelyani and Huan Liu, "Feature Selection for Classification: A Review", 2015.

[7] Dr. Vipin Saxena and Ajay Pratap, "Genetic Algorithm Based Bayesian Classification Algorithm for Object Oriented Data", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 2, no.6, December 2012, pp. 1177- 1182.

[8] Maragatham G and Lakshmi M, "Study on Classifiers using Genetic Algorithm and Class based Rules Generation", International Conference on Software and Computer Applications (ICSCA 2012), IPCSIT, 2012, Vol. 41, pp. 190 -194.

[9] Pedro Larrañaga, Hossein Karshenas a, Concha Bielza and Roberto Santana, "A review on evolutionary algorithms in Bayesian network learning and inference tasks", journal of information Sciences, Elsevier, Vol. 2, 2013, pp. 109-125.

[10] "Ensemble Methods: Foundations and Algorithms", Zhi-Hua Zhou, Chapman & Hall/Crc, Machin, CRC Press, June 6, 2012.

[11] Rajeev Singh, "Introduction to Intrusion Detection System", International Journal of Electrical and Electronics Research (IJEER) Vol. 2, Issue 1, January-March 2014, pp: 1-6.

[12] Mehdi Bahrami1 and Mohammad Bahrami, "An overview to Software Architecture in Intrusion Detection System", International Journal of Soft Computing and Software Engineering (JSCSE), Vol.1, no.1, January-March 2014, pp. 1-8.

[13] "The NSL-KDD Dataset." [Online]. Available: http://nsl.cs.unb.ca/NSL-KDD

[14] "KDD Cup 1999 Dataset." [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.