

An Analytical Comparison on Filter Feature Extraction Method in Data Mining using J48 Classifier

Abdur Rahman Onik

Department of Computer Science
and Engineering
Ahsanullah University of Science
and Technology
Dhaka, Bangladesh

Nutan Farah Haq

Department of Computer Science
and Engineering
Ahsanullah University of Science
and Technology
Dhaka, Bangladesh

Lamia Alam

Department of Computer Science
and Engineering
Military Institute of Science and
Technology
Dhaka, Bangladesh

Tauseef Ibne Mamun

Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh

ABSTRACT

The feature selection approach provides improved prediction and minimizes the computation time. Due to the higher numbers of features the understanding of the data in pattern recognition becomes difficult sometimes. That's why researchers have used different feature selection techniques with the single classifiers in their intrusion detection system to build up a model which gives a better accuracy and prediction performance. In this paper, we provide a comparative analysis with the feature selection approach in WEKA machine learning tool using the J48 classifier. The research work show the comparison of the performance of single J48 classifier with filter methods. The prediction performance may differ marginally in some cases but with the removal of irrelevant features time complexity can be easily ignored and a better prediction rate is guaranteed.

General Terms

Pattern Recognition, Data Mining.

Keywords

Intrusion Detection System, Feature Selection, Decision Tree, WEKA, Filter Method, Wrapper Method.

1. INTRODUCTION

Rapid development and the massive expansion in the computer networks causes the network security vulnerable. As the technology is rolling out, the hacking and intrusion incidents are emerging rapidly. This caused the researchers to focus on the Intrusion detection system. Intrusion detection system is assigned to protect the system from the network threats and vulnerability. Intrusion detection system can be categorized in the anomaly and misuse detection (Depren, Topallar, Anarim, & Ciliz, 2005)[4][1]. In anomaly detection, the system builds a profile of that which can be considered as normal or expected usage patterns over a period of time and triggers alarms for anything that deviates from this behavior (Adel, Zeynep & Adnan, 2014)[1]. The misuse detection is used to identify attacks in a form of signature or pattern [13]. As misuse detection uses the known pattern to detect attacks the main disadvantage is that it will fail to identify any unknown attacks to the network or system [13].

Feature selection is a part of the dimensional reduction in machine learning technique. Feature selection is a technique to reduce the irrelevant and redundant features and choose an optimal subset of features that is the representation of the dataset. Feature selection has been applied in many fields such as classification, machine learning, data mining, pattern recognition and so forth. The main objective of feature selection is to remove the redundancy and build a model with high accuracy and a better detection rate. Feature selection helps to reduce the data redundancy by removing irrelevant data that helps to reduce the time complexity of the Intrusion detection system.

Researchers have been working in the field of feature selection from early 1970 (Adel, Zeynep & Adnan, 2014), Several proposed architecture has been built with the idea of a feature selection technique.(Yinhui Li 2012) proposed an improved feature selection technique called GFR method, selected 19 features out of 41 features of NSL KDD dataset [2]. Lin, Ying, Lee, and Lee (2012) proposed simulated annealing (SA) and support vector machine (SVM) to find the best feature subsets. Simulated annealing and support vector machine were used to build a decision rule to detect new attacks [11]. (S. Devaraju, 2013) used principle component analysis for the feature selection technique. In this technique, 13 features were selected in KDD cup 99 datasets [13]. The performance of the model was measured by using the various type of neural network method. (Datti and Lakhina 2012) compared the performance of two feature reduction techniques: principal component analysis and linear discriminate analysis and used the back-propagation algorithm to test these techniques [14]. (Amin Dastanpour, 2013) used the Genetic algorithm with the Forward feature selection and Linear Correlation Feature Selection approach. Linear Correlation Feature selection selected 21 features and Forward feature selection selected 31 features out of 41 features of KDD CUP 99 dataset[3]. (Mohanabharathi, Kalaikumaran, and Karthi 2012) proposed a new feature selection method which was a combination of the information gain ratio measure and the K-means classifier [11]. Adel, Zeynep & Adnan, 2014 Proposed new feature selection approach based on the cuttlefish optimization algorithm which is used for intrusion detection systems. The proposed model uses the cuttlefish algorithm (CFA) as a search strategy to ascertain the optimal subset of features The KDD Cup 99

dataset is used to evaluate the proposed model [1]. Akhilesh, Amit, 2014 have proposed ANN-Bayesian Net-GR techniques which are an ensemble of Artificial Neural Network (ANN) and Bayesian Net with Gain Ratio (GR) feature selection technique. Both NSL-KDD and KDD cup 99 datasets being used for the testing phase [2]. Yung-Tsung, Yimeng, Tsuhan, 2010 proposed a malicious web page detection using the technique of machine learning. They analyzed the characteristic of a malicious web page systematically and presents important features for machine learning [8]. Levent Koc, Thomas, Shahram, 2012 in their proposed model categorized feature selection model in three different parts. i) Filter method ii) Wrapper method iii) Embedded method [10][1]. Ming-Yang Su, 2011 proposed a method which detect large-scale attacks, in real-time by weighted KNN [16]. They proposed a genetic algorithm combined with k nearest neighbor for feature selection and weighting [16]. Initial 35 features in the training step were weighted, along with the top ones were selected to implement testing phase [16]. For known attacks, overall accuracy rate as high as 97.42%, while only the top 19 features had been considered. While for unknown attacks, an overall accuracy rate of 78% was obtained using the top 28 features[16].

Intrusion detection system deals with a large amount of data; Feature Selection is a critical task in IDSs. In this paper, we analyzed the filter method with different search techniques and used J48 as a single classifier to build an intrusion detection model. Different filter methods with different search techniques produce an optimal subset of features. J48 is used as a single classifier to improve the quality of the produced subset of features. An analytical comparison is shown in this paper with different feature selection techniques with the single J48 classification tree. The rest of the paper is organized as follows: Section 2 presents an introduction to feature selection approaches and J48 decision tree. The comparative analysis of the feature selections techniques is discussed in section 3 highlight a brief discussion on the experimental result. Finally, Section 4 and section 5 includes the conclusion and future work part.

2. INTRODUCTION TO J48 AND FEATURE SELECTION APPROACH

2.1 Decision Tree J48

All Decision trees are most powerful technique in data mining implemented by Quinlan [2]. A decision tree offers many benefits to data mining, provides an easy understanding for the implementation and by the end user. It can proceed with flawed datasets or missing values and give a better prediction. Decision tree is capable of handling both Nominal and Numeric data. DT has three main components: nodes, arcs, and leaves. Each node splits the instance space into two or more subspaces according to a certain discrete function of the input attribute values [1]. There are two approaches in decision tree I) Univariate II) Multivariate [2]. Under the Univariate Technique, splitting is performed in the internal node using only one attribute. Such trees can be implemented using the algorithm like ID3, C4.5. In WEKA, J48 is such an extension of the C4.5 algorithm. In this paper, we work with the J48 algorithm. Quinlan's C4.5 algorithm was implemented by J48 for generating a C4.5 decision tree which can be pruned or unpruned. J48 use the concept of information gain to generate the decision tree. Under the J48 algorithm, each attribute of the dataset make a decision using the information gain and splits data into smaller subsets. The decision is made based on the highest information gain of the attribute when all the instances in the subsets belong to the same class the splitting

procedure stops. On the contrary, when features are not given any information gain, J48 can handle both continues and discrete attributes.

2.2 Feature Selection Approach

The feature selection approach removes the irrelevant features from the dataset to enhance the prediction performance and reduce the time complexity. There are three major approaches: I) embedded II) wrapper method and III) filter methods [10]. Wrapper methods use the feedback received from a specific classifier to evaluate the quality of the feature subset [10]. Wrapper method uses the predictor and it performs as a function to determine the result. A number of search methods are used which used to maximize this function that develop the performance accuracy.

Chandrashekar, Sahin, 2012 classified the wrapper method in Sequential Selection Algorithms and Heuristic Search Algorithms. According to their classification, the sequential algorithm starts with a full set and reduces the features until the objective function reached to the maximum satisfaction. An incremental criterion is introduced to fast up the procedure until the objective function reaches out the maximum satisfaction with the minimum number of features [6].

On the contrary, heuristic search algorithms evaluate different subsets to optimize the objective function. Subsets are generated either by searching around in a search space or by generating solutions to the optimization problem [6]. Filter methods evaluate on the statistical characteristics of the training data. As for its lower computational coast and reduced time complexity, this method is implemented on the large data sets like NSL KDD or KDD CUP 99 datasets. Filter method uses variable ranking techniques to reduce the irrelevant features and those ranking techniques are used due to the simplicity and their application on the practical datasets. In filter methods, the features are considered to irrelevant which are independent of the class label. There are several filter methods. In this paper, we took the two methods I) correlation- based feature selection II) consistency-based filter for the comparative analysis. The correlation-based feature selection (CFS) method ranks and selects the feature sets with biases towards subsets containing features that are highly correlated with the class and uncorrelated with each other [6]. The consistency based filter (CONS) method uses an inconsistency criterion that specifies the extent to which the dimensionally reduced data can be accepted. The algorithm generates a random subset in each round [5]. In our paper, we have done the comparative analysis using the WEKA machine learning tool. In WEKA machine learning tool, CFS Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. On the other hand, CONS evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes [6]. In our paper, we have done the comparative analysis using the WEKA machine learning tool. In WEKA machine learning tool, CFS Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. On the other hand, CONS evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes [6].

2.3 Search Techniques

Search techniques are used to identify the most irrelevant features and this technique optimizes the time complexity. WEKA uses different search techniques for the attribute selection. Under a specific attribute evaluator, several search techniques can be used. Best First Searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility. Linear Forward Selection is an Extension of Best First, taking a restricted number of k attributes into account. A fixed number k of attributes is selected by fixed-sets on the other hand, fixed-width increased k in each step. The search uses the initial ordering for selecting the top k attributes or performs a ranking. The search direction can be forward, or floating forward. Exhaustive Search performs an exhaustive search through the space of attribute subsets starting from the empty set of attributes. Greedy Stepwise performs a greedy forward or backward search through the space of attribute subsets. The search starts with all attributes or from an arbitrary point in space. Stops when the deletion of any remaining attributes results in a decrease in the evaluation. It also produces a ranked list of attributes and recording the order in which the attributes are selected. Genetic Search Performs a search using the simple genetic algorithm described in Goldberg (1989). WEKA improvises some other search techniques as well such as Race search, Ranker, Ranker Search, Random search, Scatter search, and Subset size Forward selection. In this paper, we have worked with the search techniques such as Best First, Greedy stepwise and Ranker. Ranker worked under the Info gain Attribute Evaluator. Ranker ranks attributes by their individual evaluations.

3. RESEARCH WORK And Experimental Result

In this paper, we performed the feature selection under two approaches. I) Subset Attribute Evaluator ii) Single attributes Evaluator. Under both those techniques our task was divided into two major parts, one is the attribute evaluator and the selection of the search method. We have used the J48 single classifier to classify the Intrusion detection. We have used the benchmark NSL-KDD, Breast Cancer, German Credit Card and Segment Datasets as an experimental dataset.

3.1 Dataset

The experimental dataset used in our research work is a benchmark dataset named NSL-KDD. The NSL-KDD dataset is the modified version of the KDDCUP99 dataset. Each record in the NSL-KDD dataset consists of 41 attributes and 1 class label and 4 types of attacks: Denial service attack (Dos); Remote to User (R2L); User to root (U2R) and Probe attack. The NSLKDD dataset still has some problems described by McHugh [12]. For our research, we used 30% NSL-KDD dataset as train and test set where 70% of it means 7,557 instances were used as training instance and the rest 70% means 17,634 instances were used as test instance. Figure 1 shows a detailed view of NSL-KDD data set. Table 1 gives a quick glance on the other benchmark data sets used in this research.

Type	Features
Nominal	Protocol_type(2), Service(3), Flag(4)
Binary	Land(7), logged_in(12), root_shell(14), su_attempted(15), is_host_login(21), is_guest_login(22)
Numeric	Duration(1), src_bytes(5), dst_bytes(6), wrong_fragment(8), urgent(9), hot(10), num_failed_logins(11), num_compromised(13), num_root(16), num_file_creations(17), num_shells(18), num_access_files(19), num_outbound_cmds(20), count(23) srv_count(24), error_rate(25), srv_error_rate(26), error_rate(27), srv_error_rate(28), same_srv_rate(29), diff_srv_rate(30), srv_diff_host_rate(31), dst_host_count(32), dst_host_srv_count(33), dst_host_same_srv_rate(34), dst_host_diff_srv_rate(35), dst_host_same_src_port_rate(36), dst_host_srv_diff_host_rate(37), dst_host_error_rate(38), dst_host_srv_error_rate(39), dst_host_error_rate(40), dst_host_srv_error_rate(41)

Fig 1: 41 features of NSL KDD dataset [7].

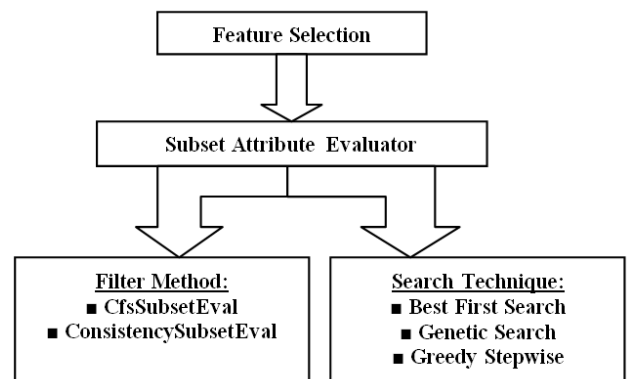


Fig 2: Feature Selection using the Subset Attribute Evaluator.

Table 1: Quick glance of other data sets used in the research.

Data Set	No. of Features	Class type	Class Labels	Instances
Breast Cancer	9	2	no-recurrence-events, recurrence-events.	286
Segment	19	7	brick-face, sky, foliage, cement, window, path, grass.	1500
German Credit Card	20	2	good, bad.	1000

3.2 Subset Attribute Evaluator:

Fig 2 demonstrates the feature selection approaches in intrusion detection system using NSL KDD dataset. Each time we selected one filter method and we run three of those search methods under that filter method. So taking those two filter method with each three search method we have a total of six

possible combinations to find out the features. We used J48 as our single classifier. The selected features using NSL-KDD data set from those combinations are given in Table 2.

Table 2: Selected features of NSL-KDD data set under each Feature selection methodology.

Feature Selection	Feature no.
CfsSubsetEval + Best First	3, 4, 5, 6, 12, 14, 26, 29, 30, 37, 38
CfsSubsetEval + GeneticSearch	2, 3, 4, 6, 10, 11, 12, 13, 15, 16
CfssubsetEval+ GreedyStepwise	3,4,5,6,12,14,26,29,30,37,38
ConsistencySubsetEval+ BestFirst	1, 3, 5, 6, 23, 32, 33, 35, 38
ConsistencySubsetEval+ GreedyStepwise	1, 3, 5, 6, 23, 32, 33, 35, 38
ConsistencySubsetEval+GeneticSearch	2,5,6,9,10,12,21,23,24,25,26, 29,32,33,35,38,41

The selected features are used to run under the J48 classifier. In each case detection accuracy is high. We can see that both Best First and Greedy stepwise select the same number of feature and selection of the features are same. Under the each feature selection the performance of the accuracy increased compared to single classifier. Though combination of the attribute evaluators with the genetic search method in terms of accuracy comparing to J48 is marginally low but the time complexity is reduced in that scenario. Fig 3 and 4 shows the performance of the each Feature selection approach using the J48 classifier. Under the CfsSubsetEval with Best First and CfsSubsetEval with Greedy stepwise the accuracy is 99.2571% and CfsSubsetEval with Genetic search is 95.0153. Attribute evaluator Consistency Subset with Best first and with Greedy stepwise search produce the same accuracy which is 99.1153%. But with Genetic search it provides a better accuracy of 99.138. The Genetic search method works proficiently with the Consistency Subset Evaluator while working with CfsSubsetEval it gives a lower accuracy rate than Single J48 classifier which accuracy is 98.064%.Table 2 enlisted the Accuracy , True positive rate and False positive rate of all feature selection approach using J48 classifier along with single J48 classifier. The Table shows that the Tp rate for CfsSubsetEval with Best First and CfsSubsetEval with Greedy Stepwise is 0.993.True positive rate for CfsSubsetEval with genetic search is 0.952. ConsistencySubsetEval with both the Best first , Genetic search and Greedy stepwise gives a true positive rate of 0.991. In term of false positive rate ConsistencySubsetEval with Best First and Greedy stepwise search gives a rate of 0.007 while in case of Genetic search it gives a rate of 0.006. False positive rate is rate is marginally better in terms of CfsSubsetEval. Attribute evaluator with the both search method Best First and Greedy stepwise is 0.006 but in case of Genetic search method the false positive rate is close to 0.003. Table 3 unlighted the Accuracy, Tp rate and FP rate of in terms of each feature selection approach.

Table 3: Comparison of Accuracy, True Positive rate, False Positive rate using J48 classifier on different Feature selection approach in NSL-KDD dataset.

Feature Selection	Accuracy	True Positive rate	False Positive rate
CfsSubsetEval + Best First	99.25	0.993	0.006
CfsSubsetEval + GeneticSearch	95.01	0.95	0.003
CfssubsetEval+ GreedyStepwise	99.25	0.993	0.006
ConsistencySubsetEval+ BestFirst	99.11	0.991	0.007
ConsistencySubsetEval+ GreedyStepwise	99.11	0.991	0.007
ConsistencySubsetEval +GeneticSearch	99.13	0.991	0.006

NSL-KDD dataset contains 4 types of attack- DoS, R2L, U2R, Probing. In our experiment, we analyzed Tp and Fp rate of each individual attacks under the selected feature selection approach with the J48 classifier. Figure 3 and Figure 4 shows the graphical representation of the accuracy rate using NSL-KDD data set. Figure 5 and Figure 6 shows the TP rate and Figure 7 and Figure 8 shows the FP rate in the graphical representation.

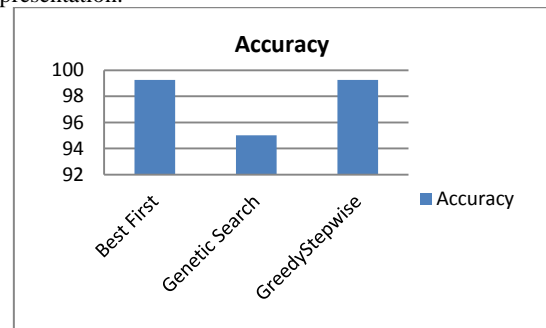


Fig 3: Accuracy rate of each search techniques of CfsSubsetEval using j48 classifier.

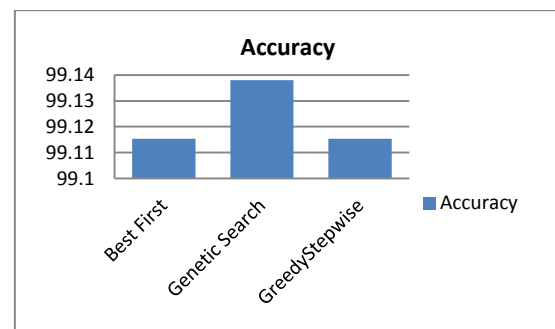


Fig 4: Accuracy rate of each search techniques of ConsistencySubsetEval using j48 classifier.

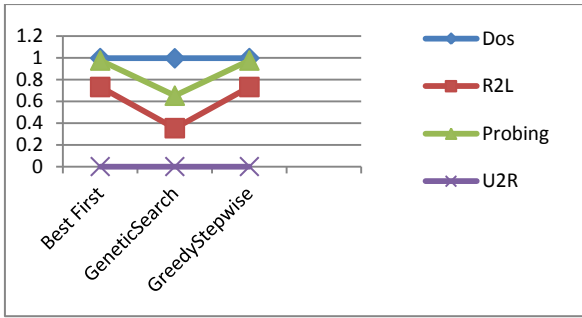


Fig 5: True Positive rate of 4 different attacks using J48 classifier under the CfsSubsetEvaluator search techniques.

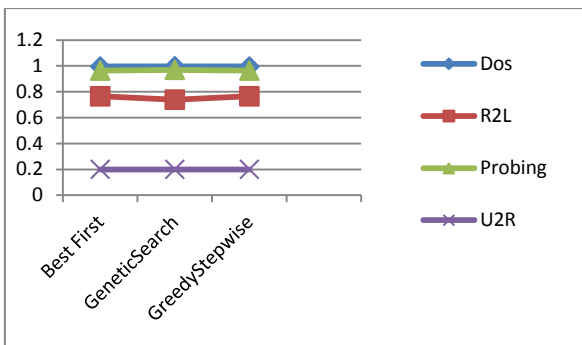


Fig 6: True Positive rate of 4 different attacks using J48 classifier under the ConsistencySubsetEval search techniques.

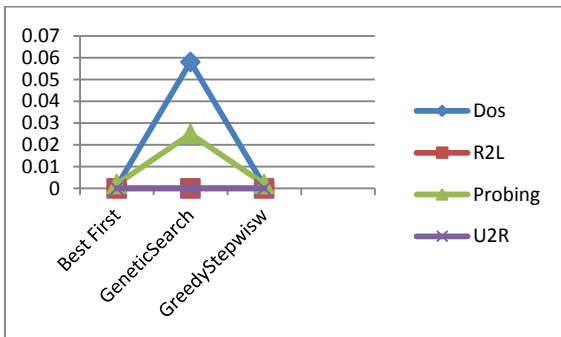


Fig 7: False Positive rate of 4 different attacks using J48 classifier under the CfsSubsetEvaluator search techniques.

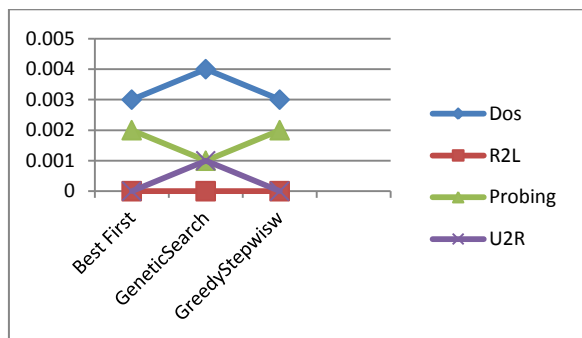


Fig 8: False Positive rate of 4 different attacks using J48 classifier under the ConsistencySubsetEval search techniques.

Fig 9 shows the graphical representation of True Positive Rate (TP) comparison of the Breast Cancer, Segment and German credit Card data set using Subset Attribute Evaluator feature selection approach. Fig 10 shows the graphical representation of False Positive Rate (FP) comparison of the Breast Cancer, Segment and German credit Card data set using Subset Attribute Evaluator feature selection approach and Figure 11 shows the accuracy comparison of the same. The detailed experimental result using Breast Cancer, German Credit Card, and Segment data set are been shown in Table 4.

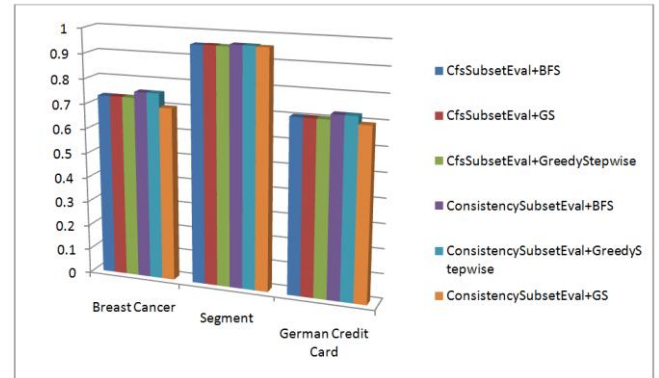


Fig 9: True Positive Rate (TP) Comparison of Breast Cancer, Segment and German credit Card data set using variety feature selection approach.

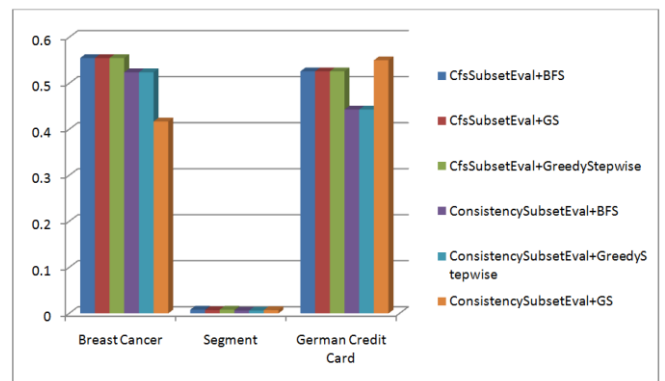


Fig 10: False Positive Rate (FP) Comparison of Breast Cancer, Segment and German credit Card data set using variety feature selection approach.

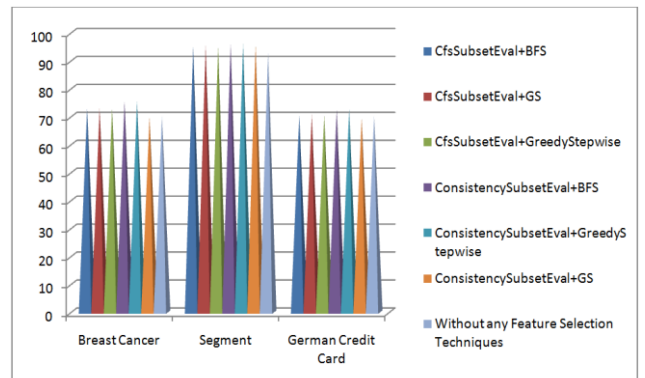


Fig 11: Accuracy Comparison of Breast Cancer, Segment and German credit Card data set using variety feature selection approach.

Table 4: Comparison of Accuracy, True Positive rate, False Positive rate using different Feature selection approach on J48 classifier in Breast Cancer, German Credit Card and Segment dataset.

Data Set	Feature Selection Approach	No. of features	Selected features	Accuracy (%)	True Positive Rate (TP)	False Positive Rate (FP)
Breast Cancer	CfsSubsetEval + Best First	5	tumor-size,inv-nodes,node-caps,deg-malig,irradiat.	73.08	0.731	0.555
	CfsSubsetEval + Genetic Search	5	tumor-size,inv-nodes,node-caps,deg-malig,irradiat.	73.08	0.731	0.555
	CfsSubsetEval + Greedy Stepwise	5	tumor-size,inv-nodes,node-caps,deg-malig,irradiat.	73.08	0.731	0.555
	ConsistencySubsetEval+ BestFirst	8	age,menopause,tumor-size,inv-nodes,deg-malig,breast,breast-quad,irradiat	75.52	0.755	0.524
	ConsistencySubsetEval+ Greedy Stepwise	8	age,menopause,tumor-size,inv-nodes,deg-malig,breast,breast-quad,irradiat	75.52	0.755	0.524
	ConsistencySubsetEval+ Genetic Search	7	age,menopause,tumor-size,deg-malig,breast,breast-quad,irradiat	70.01	0.7	0.417
Segment	CfsSubsetEval + Best First	6	region-centroid-col,region-centroid-row,hedge-mean,rawred-mean,rawgreen-mean,hue-mean.	95.33	0.953	0.008
	CfsSubsetEval + Genetic Search	8	region-centroid-col,region-centroid-row,hedge-mean,rawred-mean,rawblue-mean,exgreen-mean,saturation-mean,hue-mean.	95.4	0.954	0.007
	CfsSubsetEval + Greedy Stepwise	6	region-centroid-col,region-centroid-row,hedge-mean,rawred-mean,rawgreen-mean,hue-mean.	95.33	0.953	0.008
	ConsistencySubsetEval+ BestFirst	9	region-centroid-col,region-centroid-row,vedge-mean,hedge-mean,rawred-mean,rawblue-mean,exblue-mean,saturation-mean,hue-mean.	96.13	0.961	0.006
	ConsistencySubsetEval+ Greedy Stepwise	9	region-centroid-col,region-centroid-row,vedge-mean,hedge-mean,rawred-mean,rawblue-mean,exblue-mean,saturation-mean,hue-mean.	96.13	0.961	0.006
	ConsistencySubsetEval+ Genetic Search	9	region-centroid-col,region-centroid-row,vedge-mean,hedge-sd,intensity-mean,exred-mean,exblue-mean,saturation-mean,hue-mean.	95.8	0.958	0.007
German Credit Card	CfsSubsetEval + Best First	3	checking_status,duration,credit_history.	70.5	0.705	0.526
	CfsSubsetEval + Genetic Search	3	checking_status,duration,credit_history.	70.5	0.705	0.526
	CfsSubsetEval + Greedy Stepwise	3	checking_status,duration,credit_history.	70.5	0.705	0.526
	ConsistencySubsetEval+ BestFirst	14	checking_status,duration,credit_history,purpose,credit_amount,savings_status,employment,personal_status,other_parties,property_magnitude,age,other_payment_plans,job,own_telephone.	72.6	0.726	0.443
	ConsistencySubsetEval+ Greedy Stepwise	14	checking_status,duration,credit_history,purpose,credit_amount,savings_status,employment,personal_status,other_parties,property_magnitude,age,other_payment_plans,job,own_telephone.	72.6	0.726	0.443
	ConsistencySubsetEval+ Genetic Search	14	duration,credit_history,purpose,credit_amount,savings_status,employment,installment_commitment,personal_status,other_parties,property_magnitude,age,other_payment_plans,job,own_telephone.	69.5	0.695	0.55

3.3 Single Attribute Evaluator

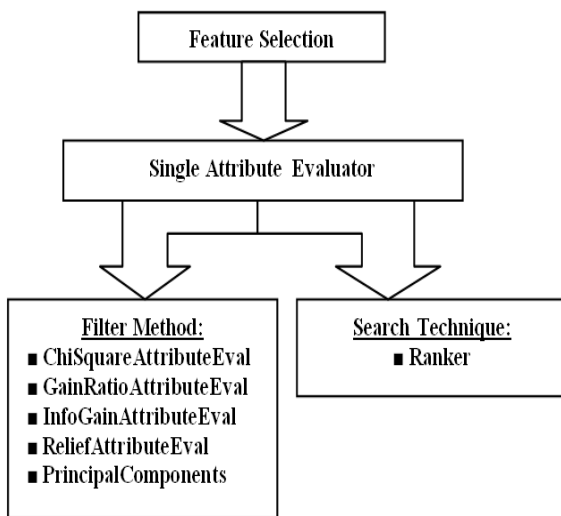


Fig 12: Feature Selection using the Single Attribute Evaluator.

Figure 12 shows the single attribute evaluating method of feature selection. ChiSquaredAttributeEval evaluates the worth of an attribute by computing the value of the chi-squared statistic with respect to the class. ReliefAttributeEval evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. It can operate on both discrete and continuous class data. GainrationAttributeEval evaluates the worth of an attribute by measuring the gain ratio with respect to the class. InfoGainAttributeEval evaluates the worth of an attribute by measuring the information gain with respect to the class. Search method Ranker ranks attributes by their individual evaluations. Under the Single Attribute Evaluator Feature selection approach determines the list of features according to the ranking. The ranking order shows the importance of each feature in the determination of the class label correctly. Table 5 shows the Ranked attributes of each feature selection techniques.

4. CONCLUSION

In data mining, the decision tree is very much popular in use and for that reason in this research we tried to find out a way for improving the performance of J48 classifier by reducing redundant features. As the filter feature selection method is independent of the classifier, so we conduct our experiment by evolving through the various filter methods and analysis the performance by using J48 as the classification algorithm. In our experimental analysis, we found out that most of the time CfsSubsetEval produces lesser important features than the ConsistencySubsetEval. Using NSL-KDD data set we got maximum 99.25% of accuracy by the J48 classifier using only 11 important features within 41 features extracted by CfsSubsetEval and Greedy Stepwise search technique. Where the same classifier shows an accuracy of 98.33% using all 41 features. When we have conducted the same experiment using Breast Cancer dataset, at first it showed an accuracy of 70.01% using all 9 features but the accuracy increased to 75.52% using only 8 important features extracted by ConsistencySubsetEval and both the Best First along with the Greedy Stepwise search techniques. Again, by using

ConsistenctSubsetEval feature extraction methodology with Best First search technique, extracted the best feature set which achieved the maximum accuracy of 72.6% using J48 classifier in German Credit Card data set. The same result is gained by using Greedy Stepwise search technique instead of Best First. The same feature extraction methods helps to get the maximum accuracy of 96.13% while using Segment data set by extracting only 9 features within 19 features of it; where using all the 19 features it only achieved an accuracy of 94.7%. So, from all our experiments it is been clearly seen that ConsistencySubsetEval feature extraction methodology with Best First or Greedy Stepwise search technique gives us a better performance most of the time. Our experiment was conducted by using 4 benchmark data sets for showing the experimental results are not in dependency upon data sets.

5. FUTURE WORK

In this paper, we have conducted our experiments in the path of filter feature extraction search methods. We have evolved in most of the part of filter method though some of them are yet to be explored. In this paper, we have worked only with the offline data sets. So our future work will be focused on:

- Exploring other feature selection techniques like, wrapper approach or embedded one.
- Exploring feature extraction techniques in other classifiers.
- Analysis the existing feature extraction methods and work on building up a new one.
- Along with the offline datasets, in future we will try to work with the online data set and analysis the result.

6. REFERENCES

- [1] Adel Sabry Eesa , Zeynep and Brifcani (2015). A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems in Expert Systems with Applications. Volume 42, Issue 5, Pages 2670-2679.
- [2] Akhilesh Kumar Shrivastava, Amit Kumar D, (2014). "An Ensemble Model for Classification of Attacks with Feature Selection based on KDD99 and NSL-KDD Data Set." *International Journal of Computer Applications*.
- [3] Amin Dastanpour, Raja Azlina, (2013). "Feature Selection Based on Genetic Algorithm and Support Vector Machine for Intrusion Detection System." *The Society of Digital Information and Wireless Communications (SDIWC)*.
- [4] Depren, O., Topallar, M., Anarim, E., & Ciliz, M. K. (2005). An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks. *Expert Systems with Applications*, 29(4), 713–722.
- [5] Dr. Neeraj Bhargava, Girja Sharma and Dr. Ritu Bhargava,(2013),"Decision Tree Analysis on J48 Algorithm for Data Mining", *IJARCSSE*, Volume 3, Issue 6, June 2013 .
- [6] Girish Chandrashekar, Ferat Sahin, (2014),A survey on feature selection methods in *Computers and Electrical Engineering archive*, Volume 40, Issue 1, January, 2014 , Pages 16-28.
- [7] Hee-su Chae, Byung-oh Jo,Sang-Hyun Choi and Twae-kyung Park (2015) , "Feature Selection for Intrusion Detection using NSL-KDD", *Recent Advances in Computer Science*, ISBN: 978-960-474-354-4.

[8] Hou, Yung-Tsung, et al. "Malicious web content detection by machine learning." *Expert Systems with Applications* 37.1 (2010): 55-60.

[9] Law MH, Figueiredo M rio AT, Jain AK. (2004). Simultaneous feature selection and clustering using mixture models in *IEEE Trans Pattern Anal Mach Intell*, Volume 26, Issue 9, September, 2004, Pages :1154-66.

[10] Levent Koc, Thomas A. Mazzuchi and Shahram Sarkani,(2012),A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier in *Expert Systems with Applications* , Volume 39, Issue 18, Pages 13492-13500.

[11] Lin, S.-W., Ying, K.-C., Lee, C.-Y., & Lee, Z.-J. (2012). An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection. *Applied Soft Computing*, 12(10), 3285–3290.

[12] Mohanabharathi, R., Kalaikumar, T., & Karthi, S. (2012). Feature selection for wireless intrusion detection system using filter and wrapper model in *International Journal of Modern Engineering Research (IJMER)*, 2(4), 1552–1556.

[13] Nutan Farah Haq, Abdur Rahman Onik, (2015). "Application of Machine Learning Approaches in Intrusion Detection System: A Survey." (*IJARAI*)

[14] Rupali Datti, Shilpa Lakhina (2012), "Performance Comparison of Feature Reduction Techniques For Intrusion Detection Systems", Dated: 10-02-2012, *International Journal of Computer Science and Technology, IJCST*, ISSN : 0976-8491, Volume 3, Issue 1.

[15] S. Devaraju, S. Ramakrishnan[2013]. "DETECTION OF ACCURACY FOR INTRUSION DETECTION SYSTEM USING NEURAL NETWORK CLASSIFIER." *International Journal of Emerging Technology and Advanced Engineering(IJETAE)*.

[16] Su, Ming-Yang, (2011). "Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers." *Expert Systems with Applications* 38.4 : 3492-3498.

[17] Yang Yi, Jiansheng Wu (2011). "Incremental SVM based on reserved set for network intrusion detection." *expert systems with applications,ELSEVIER*.

[18] Yinhui Li, Jingbo Xia (2012). "An efficient intrusion detection system based on support vector machines and." *expert systems with applications,ELSEVIER*.

7. APPENDIX

Table 5: Selected features of NSL-KDD data set under each Feature selection methodology in Single Attribute Evaluator.

Feature selection approach (Attribute evaluator + Search method)	Ranking of the attributes
ChiSqrAttributeEval + Ranker	5,3,6,35,4,30,33,29,34,23,25,38,39,37,26,36,12,32,10,24,40,31,27,2,41,1,28,9,14,22,13,17,11,8,16,19,15, 18,20,7,21
GainrationAttributeEval + Ranker	9,26,25,4,12,39,30,38,6,29,5,37,11,3,22,35,34,14,33,23,8,10,31,27,28,32,1,36,2,41,40,17,13,16,19,18,15, 24,7,20,21
InfoGainAttributeEval + Ranker	5,3,30,4,6,29,35,23,33,34,38,25,39,26,12,37,36,32,24,31,40,2,41,1,28,10,22,8,13,16,14,17,19,11,15,9,18, 7,20,21
ReliefAttributeEval + Ranker	3,29,4,36,32,38,12,33,2,34,39,23,26,35,40,30,31,8,24,25,37,27,41,28,10,22,1,6,14,11,13,15,19,18,16,5,17, 9,7,20,21
PrincipaleComponents + Ranker	3, 41,28,10,22 ,12,33,2,34,39,23,26,35,40,30,31,8,24,25,37,27 ,1,6,14,29,4,36,32,38,11,13,15,19, 18,16,5,17,9,7,20,21