# Performance Evaluation of Feature Selection Algorithms in Educational Data Mining

**Article** · December 2016

DOI: 10.20894/IJDMTA.102.005.002.007



Velmurugan T
D.G.Vaishnav College

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.131-139
ISSN: 2278-2419

# Performance Evaluation of Feature Selection Algorithms in Educational Data Mining

T.Velmurugan[1], C. Anuradha[2]
[1]Associate Professor, PG and Research Dept. of Computer Science, D.G.Vaishnav College, Chennai, India.
[2]Research Scholar, BharathiarUniversity, Coimbatore, India.
Email: velmurugan_dgvc@yahoo.co.in;anumphil14@gmail.com;

**Abstract:** Educational Data mining(EDM)is a prominent field concerned with developing methods for exploring the unique and increasingly large scale data that come from educational settings and using those methods to better understand students in which they learn. It has been proved in various studies and by the previous study by the authors that data mining techniques find widespread applications in the educational decision making process for improving the performance of students in higher educational institutions. Classification techniques assumes significant importance in the machine learning tasks and are mostly employed in the prediction related problems. In machine learning problems, feature selection techniques are used to reduce the attributes of the class variables by removing the redundant and irrelevant features from the dataset. The aim of this research work is to compares the performance of various feature selection techniques is done using WEKA tool in the prediction of students' performance in the final semester examination using different classification algorithms. Particularly J48, Naïve Bayes, Bayes Net, IBk, OneR, and JRip are used in this research work. The dataset for the study were collected from the student's performance report of a private college in Tamil Nadu state of India. The effectiveness of various feature selection algorithms was compared with six classifiers and the results are discussed. The results of this study shows that the accuracy of IBK is 99.680% which is found to be high than other classifiers over the CFS subset evaluator. Also found that overall accuracy of CFS subset evaluator seems to be high than other feature selection algorithms. The future work will concentrate on the implementation of a proposed hybrid method by considering large dataset collected from many institutions.

Keywords: Educational data mining, Wrapper selection, Best First Search,Classification Algorithms, Feature selection Algorithms.

## I. INTRODUCTION

In educational data mining, prediction of students' performance has long been an interesting area of research and it helps to identify the weak students or students at risk. As the educational institutions are facing intense competition with respect to admission, retention and sustenance, it is important that the institutions pay significant attention in improving students output. Most often the institutions are judged by the percentage of results produced by students' in the finale end semester examination. A data mining system offer several techniques to the educational leaders to support the decision making process to improve the quality of education. The large volume of data generated in the educational institutions can be effectively used to draw rich source of vital information to support decision making system. The main focus of this research work is to identify the best feature selection and classification algorithms to examine a performance of undergraduate student performance in education data set. The objective is to find the best attribute by comparing the performance of various feature selection techniques in the prediction of students' performance in the final semester examination using different classification algorithms such as J48, Naïve Bayes, Bayes Net, IBk, OneR, and JRip are used in this research work. The idea behind this research work is to identify slow learners which help the faculties to give special attention to individual student's to improve their academic performance.

A research work done by parneet kaur et al. in education sector. Their work focuses on identifying the slow learners among students and applies feature selection algorithms to filter desired potential variables using WEKA tool. As a result, statistics are generated based on all classification algorithms in order to predict the accuracy [*]. Another work by HythemHashim et al. discussed about Data mining methodologies to study student's academic performance using the C4.5 Algorithm. Their objective is to build a classification model that can be used to improve the student's academic records in Faculty of Mathematical Science and Statistics. This model has been done using C4.5 for predicting student performance in many different settings [1].

A work done by Vaibhav and Rajendra named as Classification and performance evaluation using data mining algorithms. The authors collected student data from polytechnique institute and classified the data using Decision tree and Naïve Bayesian algorithms. They compare results of classification with respect to different performance parameters [2]. Another research done by Anjana and Jeena discussed about Predicting College Students Dropout using EDM Techniques. Here WEKA tool has been used to evaluate the attributes. Various classification techniques like induction rules and decision tree have been applied to data and results of each of these approaches have been compared [3]. A paper Titled "Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue" by Pooja et al. has been done towards the usage of data mining techniques in the field of education. This paper presents a comprehensive survey towards educational data mining [4]. A work by Punlumjeak and Rachburee had proposed a comparison of feature selection techniques namely genetic algorithms, support vector machine, information gain, minimum and maximum relevance algorithms with supervised classifiers such as naïve bayes, decision tree, k-nearest neighbour and neural network. Their results shows that minimum and maximum relevance feature selection method with 10 features give the best result on

Volume: 05 Issue: 02 December 2016 Page No.131-139
ISSN: 2278-2419

91.12% accuracy with k-nearest neighbour classifier[5]. Another work by Anal and Devadatta had applied a different feature selection algorithm on the student data set. The best results are achieved by correlation based feature selection with 8 features. Subsequently classification algorithms may be applied on this feature subset for predicting student grades [6].Komal and Supriya [7] have conducted a Survey on Mining Educational Data to Forecast Failure of Engineering Students. This paper provides a Review of the available literature on Educational Data mining, Classification method and different feature selection techniques that author should apply on Student dataset. The research paper titled Improvement on Classification Models of Multiple Classes through Effectual Processes by Tarik [8].

This paper work focuses on improving the results of classification models of multiple classes via some effective techniques. The collected data are pre-processed, cleaned, filtered, normalized, the final data was balanced and randomized, then a combining technique of Naïve Base Classifier and Best First Search algorithms are used to ultimately reduce the number of features in data sets. Finally, a multi-classification task is conducted through some effective classifiers such as K-Nearest Neighbor, Radial Basis Function, and Artificial Neural Network to forecast the students' performance. Another work carried out by Sadaf and Kulkarni discussed about Precognition of Students Academic Failure Using Data Mining Techniques. This research paper proposes to pre-recognize Student's academic failure using various Data mining techniques especially induction rules, decision trees and naïve Bayes are applied [9].

Carlos et al. [10] have tried to attempt to solve this problem of predicting student's academic failure using clustering algorithms, induction rules or decision trees algorithms of data mining techniques. Authors applied five rules of induction rules and five decision tree algorithms on the dataset. Sahil and Shweta have carried out a Study of Application of Data Mining and Analytics in Education Domain. This paper basically is a study of certain research experiments which aims to study the different applications of data mining techniques on the educational data. Also it elaborated upon the state of the art techniques in the field of educational analytics [11]. Ogunde and Ajibade have developed a new system for the prediction of students graduation grades based on entry results data. The proposed system uses ID3 algorithm to classify the data and construct the decision tree by employing a top-down, greedy search to test every attributes [12].Dinesh and Radika had done a survey on predicting Student academic Performance in educational environment which is based upon the psychological and environmental factor is predicted using different educational data mining techniques. Researchers also survey the predictive model in data mining and current trends in prediction in data mining [13].

A Work done by ArpitTrivedi has put forward a simple approach for categorizing student data using decision tree based approach. For taking measures of category of specific student, a frequency measure is used as a feature extraction. With the use of trained classifier, they predicted the class for indefinite student automatically [14].A work has done by Agrawal and Gurav have done a review on Data Mining Techniques Used for Educational System. This paper is based on survey which proposes to apply data mining techniques such as association rule mining, classification techniques [15].The classification is a data mining technique which includes systematic approach to building the classification models from an input dataset [16]. Some of the popular classifiers used to solve a classification problem are decision tree classifiers, rule-based classifiers, neural networks, support vector machines, and naive Bayes classifiers [17]. Therefore, a key objective of the learning algorithm is to build a predictive model that accurately predicts the class labels of previously unknown records. This paper examines that various classification algorithms and their performance are compared using WEKA software and results are discussed. The open source data mining tool WEKA was used in the present work to obtain the reduced set of features from the available feature set using various feature selection techniques. In addition, the reduced attributes were given as input to the classifiers like decision tree algorithm C4.5 (J48), Bayesian classifiers like Naïve Bayes and BayesNet, Nearest Neighbor algorithm (IBk) and rule learners (OneR and JRip) to evaluate the performance of the classification algorithms for the particular feature selection technique.

This paper is structured as follows. Section 2 discusses about background of the study. Section 3 describes various feature selection techniques used for reducing the attributes of the dataset. The statement of the problem is provided in Section 4. The details of the dataset generated for the study is presented in the Section 5. The experimental evaluation and comparative analysis are given in Section 6 and Conclusion for the proposed work is given in Section 7. Finally, vital references are mentioned in Section 8.

## II. BACKGROUND

Feature selection has been an important field of research in data mining and machine learning systems. The primary objective of any feature selection technique is to choose a subset of features of the input variables by eliminating those features which are redundant, irrelevant orof no predictive information [18]. Feature subset selection in machine learning can be broadly classified into three groups as filter, wrapper and embedded models [19]. Filters based method of feature selection depends on the general characteristics of the training data. Thus, feature selection process is carried out as a pre-processing step, independent of the learning algorithm. Wrapper technique depends on the learning algorithm and uses it as a black box to evaluate the usefulness of subsets of variables in the prediction task. Thus, wrapper methods uses learning algorithm to evaluate the subset of features for feature selection. Wrapper methods are computationally intensive. Embedded methods on the other hand perform feature selection during the training process of the classifier. This methods are particularly specific to a given learning machines.

As the dimensionality of a domain expands, the number of features N increases. Finding an optimal feature subset is intractable and problems related feature selections have been proved to be NP-hard. At this juncture, it is essential to describe traditional feature selection process, which consists of four basic steps, namely, subset generation, subset evaluation, stopping criterion, and validation. Subset generation is a search process that produces candidate feature subsets for evaluation

Volume: 05 Issue: 02 December 2016 Page No.131-139
ISSN: 2278-2419

based on a certain search strategy. Each candidate subset is evaluated and compared with the previous best one according to a certain evaluation. If the new subset turns to be better, it replaces best one. This process is repeated until a given stopping condition is satisfied [20]. A number of studies have established in theory and practice that feature selection is an effective technique in improving learning efficiency, enhancing predictive accuracy and minimizing the complexity of results in data mining system. The effectiveness of feature selection has been proved in many applications involving data mining and machine learning like text categorization [21], image retrieval [22], information retrieval [23], DNA microarray analysis [24], intrusion detection [25,26], and music information retrieval [27].

## III. STATEMENT OF THE PROBLEM

In this research work, the performance of various feature selection algorithms was evaluated on different classification algorithm using the students' academic performance dataset generated for the study. The proposed study made several comparisons to evaluate the effectiveness of the feature selection techniques using the measures involving error and accuracy parameters. The overall aim of the study was to analyze the effectiveness of various machine learning algorithms to predict students' performance in the end semester examination. The dataset for the study included the demographic details of the students like gender, family size and type, income, parent's educational attainment and locality. In addition, pre-collegiate conditions of the students like their performance in secondary and higher secondary classes are also collected and maintained in the colleges. Thus, it could be useful to the educational leaders and management of the colleges, if the features in the currently available data can be acting as the indicator for predicting the performance of the students. The major objective of this study is to analyze the student's data available in the degree colleges to identify any specific patterns that might be useful in the prediction of their performance in the university exams. The specific objective of the study is to classify students according to their performance in the final examination based on their personal and pre-collegiate characteristics.

## IV. RESEARCH METHODOLOGY

In this research work six classification algorithm are used such as J48, Naïve bayes, Bayes net, IBK, OneR and JRip along with four feature selection algorithms. In this section, the fundamentals of some the feature selection algorithms are illustrated. Furthermore, the algorithms CfsSubset evaluations, Chi-Squared Attribute Evaluation, Information Gain Attribute Evaluation and Relief attribute evaluation which are used in this research work are also described.

### A. Correlation-based Feature Selection (CFS)

Correlation based Feature Selection (CFS) is a simple filter algorithm that ranks feature subsets according to a correlation based heuristic evaluation function [28]. In CFS, the bias of the evaluation function is toward subsets that contain features that are highly correlated with the output to be predicted and uncorrelated with each other. Irrelevant features should be ignored because they will have low correlation with the class. Redundant features should be screened out as they will be highly correlated with one or more of the remaining features. The acceptance of a feature will depend on the extent to which it predicts classes in areas of the instance space not already predicted by other features.

### B. Best First Search Algorithm (BFS)

Best First Search is an approach that searches the attribute subsets space via a method of Greedy Hill Climbing improved with a backtracking aptitude. The controls of the amount of backtracking can be achieved via setting the quantity of consecutive non-improving nodes. This approach might start to search in both directions; forwardly or backwardly. It can start with the empty set of attributes and search forwardly, or it can start with the full set of attributes and search backwardly [8].The Table 1 shows the Best first search algorithm [28].

Table 1: Best first search algorithm

| |
|---|
| 1. Begin with the OPEN list containing the start state, the CLOSED list empty, and BEST← start state. |
| 2. Let s = arg max e(x) (get the state from OPEN with the highest evaluation). |
| 3. Remove s from OPEN and add to CLOSED. |
| 4. If e(s) _ e (BEST), then BEST ← s. |
| 5. For each child t of s that is not in the OPEN or CLOSED list, evaluate and add to OPEN. |
| 6. If BEST changed in the last set of expansions, goto 2. |
| 7. Return BEST. |

### C. Wrapper Feature Selection

In the wrapper approach, the feature subset selection is done using the induction algorithm as a black box. The feature subset selection algorithm conducts a search for a good subset using the induction algorithm itself as part of the evaluation function. The accuracy of the induced classifiers is estimated using accuracy estimation techniques. Wrappers are based on hypothesis. They assign some values to weight vectors, and compare the performance of a learning algorithm with different weight vector. In wrapper method, the weights of features are determined by how well the specific feature settings perform in classification learning. The algorithm iteratively adjusts feature weights based on its performance [29].

### D. CfsSubset Evaluator (CSER)

It 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 [7].

### E. Chi-Squared Attribute Evaluator (CSAER)

ChiSquaredAttributeEval evaluates an attribute by computing the value of the chi-squared statistic with respect to the class [7].

### F. Information Gain Attribute Evaluator (IGAER)

It Evaluates an attribute by measuring the information gain with respect to the class Info Gain (Class, Attribute) = H (Class) - H (Class | Attribute) [7].

### G. Relief Attribute Evaluator (RAER)

It 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 can operate on both discrete and continuous class data [7].

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.131-139
ISSN: 2278-2419

Table 2: Description of options and Capability of CfsSubset evaluator

| **Option** | **Description** |
| --- | --- |
| locally Predictive | Identify locally predictive attributes. Iteratively adds attributes with the highest mutual relationship with the class as long as there is not already an attribute in the subset that has a higher correlation with the attribute in question |
| missingSeperate | Take missing as a separate value. |
| **Capability** | **Supported** |
| Class | Missing class values, Numeric class, nominal class, Date class, Binary class |
| Attributes | Empty nominal attributed, Nominal attributes, Numeric Attributes, Unary attributes, Date attributed, Binary attributes, Missing Values |
| Min # of instances | 1 |

Table 3: Description of options and Capability of Chi-Squared Attribute Evaluator

| **Option** | **Description** |
| --- | --- |
| BinarizeNumericAttributes | Only binarize numeric attributes instead of properly discretizing them. |
| Missing Merge | Distribute the counts for missing values. Then counts are distributed across other values in proportion to their frequency. Or else, missing is treated as a separate value. |
| **Capability** | **Supported** |
| Class | Missing class values, nominal class, Binary class |
| Attributes | Empty nominal attributed, Nominal attributes, Numeric Attributes, Unary attributes, Date attributed, Binary attributes, Missing Values |
| Min # of instances | 1 |

Table 4: Description of options and Capability of Info Gain Attribute Evaluation

| Option | Description |
| --- | --- |
| binarizeNumericAttributes | Just binarize numeric attributes rather than properly discretizing them |
| Missing Merge | Distribute the counts for missing values. Counts are distributed over other values in proportion to their frequency. Or else, missing is treated as a separate value. |
| Capability | Supported |
| Class | Missing class values, nominal class, Binary class |
| Attributes | Empty nominal attributed, Nominal attributes, Numeric Attributes, Unary attributes, Date attributed, Binary attributes, Missing Values |
| Min # of instances | 1 |

Table 5: Description of options and Capability of Relief Attribute Evaluation

| **Option** | **Description** |
| --- | --- |
| numNeighbors | Number of nearest neighbors for attribute estimation |
| sample Size | Number of instances to sample. Default (-1) indicates that all instances will be used for attribute estimation. |
| Seed | Random seed for sampling instances |
| Sigma | Set influence of nearest neighbors. Used in an exp function to control how quickly weights decrease for more distant instances. Use in conjunction with weight By Distance. Sensible values = 1/5 to 1/10 the number of nearest neighbors. |
| WeightByDistance | Weight nearest neighbors by their distance |
| **Capability** | **Supported** |
| Class | Nominal class, Date class, Missing class values, numericclass, Binary class |
| Attributes | Empty nominal attributes, Nominal attributes, Numeric Attributes, Unary attributes, Date attributed, Binary attributes, Missing Values |
| Min # of instances | 1 |

## V. EXPERIMENTAL DATA

A student’s dataset was generated based on the demographic characteristics, student’s admission dataand pre-collegiate features of the students. In addition, performance related measures were also gathered based on class and university examinations. The study data mining classification algorithms that are compared in the study includes Naive Bayes, Bayes Net Classifiers [30], and OneR, J48 decision tree algorithm which is an open source Java implementation of C4.5 algorithm [31], IBK, JRip algorithm [32] and J48 algorithms. The information used in this study was collected from college students enrolled in Bachelor degree program at a 3 reputed arts and Science College in the state of Tamil Nadu affiliated to Thiruvalluvar University in the year 2014.The total number of student’s data was 610 students with 21 attributes were collected through questionnaire. The collected data was organized in Microsoft Excel sheet.

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.131-139
ISSN: 2278-2419

The target variable was Student End Semester Marks (ESM) which was usually in numeric form in terms of percentage. It was discretized using pre-processing filters into 4 categories. The categories of target variable included First Class (Score > 60%), Second Class (45 - 60%), Third Class (36 - 45%), Fail (< 36%). Each student record had the following attributes based on student personal data included gender, category of admission, living location, family size, and family type, annual income of the family, father's qualification and mother's qualification. The attributes referring to the students' pre-college characteristics includedStudents Grade in High School and Students Grade in Senior Secondary School. The attributes describing other college features include thebranch of study of the students, place of stay, previous semester mark, class test performance, seminar performance, assignment, general proficiency, class attendance and performance in the laboratory work. The study was limited to student's data collected from three Arts and Science Colleges in Tamil Nadu. The detailed description of the dataset is provided in Table 6.

Table 6: Description of the attributes used for Classification

| **Variables** | **Description** | **Possible Values** |
|---|---|---|
| Gender | Students Sex | {Male, Female} |
| Branch | Students Branch | {BCA, B.SC, B.COM, B.A} |
| Cat | Students category | {BC, MBC, MSC, OC, SBC, SC} |
| HSG | Students grade in High School | {O – 90% -100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 35% - 49%, FAIL - <35%} |
| SSG | Students grade in Senior Secondary | {O – 90% -100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 35% - 49%, FAIL - <35% } |
| Medium | Medium of instruction | Tamil, English, others |
| LLoc | Living Location of Student | {Village, Taluk, Rural, Town, District} |
| HOS | Student stay in hostel or not | {Yes, No} |
| FSize | student's family size | {1, 2, 3, >3} |
| FType | Students family type | {Joint, Individual} |
| FINC | Family annual income | {poor, medium, high} |
| FQual | Fathers qualification | {no-education, elementary, secondary, UG, PG, Ph.D} |
| MQual | Mother's Qualification | {no-education, elementary, secondary, UG, PG, Ph.D. NA} |
| PSM | Previous Semester Mark | {First > 60%, Second >45 &<60%, Third >36 &<45% Fail < 36%} |
| CTG | Class Test Grade | {Poor, Average, Good} |
| SEM_P | Seminar Performance | {Poor , Average, Good} |
| ASS | Assignment | {Yes, No} |
| GP | General Proficiency | {Yes, No} |
| ATT | Attendance | {Poor , Average, Good} |
| LW | Lab Work | {Yes, No} |
| ESM | End Semester Marks | {First > 60% , Second >45 &<60% , Third >36 &<45%, Fail < 36%} |

## VI. EXPERIMENTAL SETUP

The main objective of this research is to study the impact of feature selection techniques on the classification task so that classification performance can be improved in the prediction of student performance for the student performance dataset generated in the study. The classification model was built using different algorithms like Naive Bayes, BayesNet, OneR, IBK, JRip and J48. The WEKA application was used for this purpose. Each classifier is applied for two testing options - cross validation (using 10 folds and applying the algorithm 10 times - each time 9 of the folds are used for training and 1 fold is used for testing) and percentage split (2/3 of the dataset used for training and 1/3 – for testing).The Feature selection algorithm tries to select those attributes which have greater impact on their academic status. Feature Selection Algorithms used in this study are as follows: CfsSubsetEval, Chi-SquaredAttributeEval, InfoGainAttributeEval, and ReliefAttributeEval. Table 7 shows the best attributes that have selected by Feature Selection Algorithms using WEKA software tool.

Table 7: Reduction of Attributes using Feature selection algorithm

| **Feature subset Algorithm** | **Attributes** | **No. of Attributes** |
|---|---|---|
| Without Feature Selection Algorithms | Sex, Branch, Cat, SSG_Grade, HSG_Grade, Medium, LOC, HOS, FSIZE, FTYPE, FINC, FQUAL, MQUAL, PSM, CTG, SEM_P, ASS, GP, ATT, LW, ESM | 21 |
| CSER | Branch, SSG_Grade, FINC, PSM, GP, ATT, ESM | 7 |
| CSAER | PSM, Branch, FINC, ATT, SSG_Grade, LW, FSTAT, CTG, Medium, GP, Sex, ESM | 12 |
| IGAER | PSM, Branch, FINC, LW, ATT, FSTAT, SSG_Grade, Medium, | 10 |

| | | |
|---|---|---|
| | CTG, ESM | |
| RAER | Branch, PSM, FINC, LW, Medium, CTG, FSTAT, ESM | 8 |

## VII. RESULTS AND DISCUSSIONS

The Performance of this model is highly depends on selection of best Attributes from the list of attribute used in student data set used in student data set. The present investigation focuses on different Feature Selection Algorithm used in data preprocessing. Effectiveness of the algorithm is presented in terms of different measures. For assessing the goodness here Receiver Operating Characteristics (ROC) value can be used.ROC value is the representation of the tradeoff between the false positive and false negative rates. F-Measure, which is another measure for evaluating the effectiveness, is the harmonic mean of the precision and recall. The evaluation measures with variations of ROC values and F-Measure are generated from an Open Source Data mining tool WEKA.

Average F-measure was computed for feature selection techniques for each of the classification algorithms. F-measure determines the predictive accuracy of the classifier. Ten-fold cross validation was used for this purpose. The results are summarized in Table 8. The average F-measure was also calculated for without performing feature selection. The results clearly show that CSER (0.972) has outperformed other techniques using 7 attributes. RAER has also produced better F-Measure (0.961) with 8 attributes.

Table 8: Average F-Measure for each feature subset

| Feature Subset | F – Measure | | | | | | Average F-Measure |
|---|---|---|---|---|---|---|---|
| | Naive Bayes | Bayes Net | One R | IBK | JRip | J48 | |
| Without feature Selection | 0.943 | 0.924 | 0.915 | 0.953 | 0.939 | 0.944 | 0.936 |
| CSER | 0.996 | 0.969 | 0.980 | 0.997 | 0.984 | 0.909 | 0.972 |
| CSAER | 0.956 | 0.925 | 0.983 | 0.956 | 0.971 | 0.85 | 0.940 |
| IGAER | 0.959 | 0.936 | 0.951 | 0.952 | 0.946 | 0.907 | 0.941 |
| RAER | 0.958 | 0.923 | 0.981 | 0.97 | 0.962 | 0.976 | 0.961 |

A. Without Feature Selection Algorithm (21 Attributes)

In the present study, Classifiers was implemented Without Feature Selection algorithm (WFS) applied on the data set and results of the classifiers shown in Table 9. The results reveal that the True Positive rate is high for the IBK and OneR, while it is low for the classifier Bayes Net. The Precision is high for OneR classifier and it is very low the classifier IBK.

Table 9: Classification results for WFS

| Classifier | TP Rate | Precision | F-Measure | ROC Area |
|---|---|---|---|---|
| J48 | 0.95 | 0.95 | 0.944 | 0.969 |
| Naïve Bayes | 0.98 | 0.03 | 0.943 | 0.998 |
| Bayes Net | 0.94 | 0.953 | 0.924 | 0.987 |
| IBk | 0.985 | 0.029 | 0.953 | 0.98 |
| OneR | 0.985 | 0.985 | 0.915 | 0.971 |
| JRip | 0.97 | 0.971 | 0.939 | 0.745 |

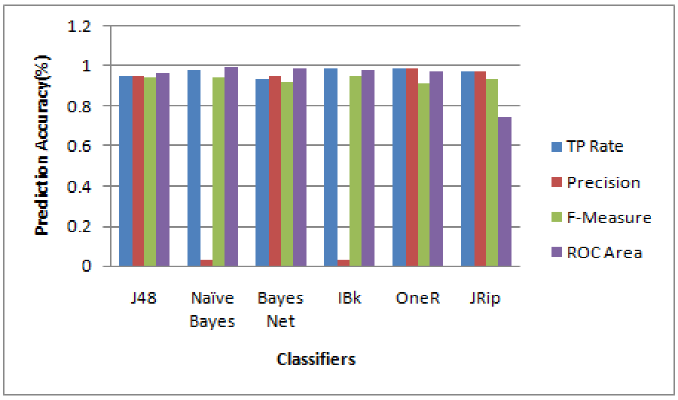


Figure 1: Results of WFS

B. CfsSubsetEval with Best First Search Algorithm (7 attributes)

The present study implements CfsSubset Evaluator has been implemented and results are presented in Table 10. The Table presents that IBK classifier correctly classifies about 99.680% for the 10-fold cross-validation testing. It also shows that the True positive rate is high for Bayes Net. The Precision value is high for Naïve Bayes and low for J48.

Table 10: Classification results for the CfsSubset Evaluator

| Classifier | TP Rate | Precision | Recall | F-Measure | ROC Area |
|---|---|---|---|---|---|
| J48 | 0.965 | 0.96 | 0.965 | 0.909 | 0.968 |
| Naïve Bayes | 0.986 | 0.998 | 0.996 | 0.996 | 0.986 |
| Bayes Net | 0.997 | 0.996 | 0.967 | 0.969 | 0.987 |
| IBk | 0.99 | 0.99 | 0.99 | 0.997 | 0.984 |
| OneR | 0.980 | 0.980 | 0.980 | 0.980 | 0.983 |
| JRip | 0.985 | 0.985 | 0.985 | 0.984 | 0.987 |

From the Fig.3, we observe that the classifier Bayes Net and JRip has the highest ROC value of 0.987 when it had 7 attributes. The generated macro-averaged F-measure could attain a maximum of 0.997 for the classifier IBk.

C. Chi-SquareAttributeEval and Ranker (12 Attributes)

The present study implements Chi-Square Attribute Evaluator has presented in Table 11. The results from Table 6 reveal that the True Positive rate and precision is high for OneR classifier and low for J48. It can be verified that OneR classifier

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.131-139
ISSN: 2278-2419

correctly classifies about 98.314% for the 10-fold cross-validation testing.

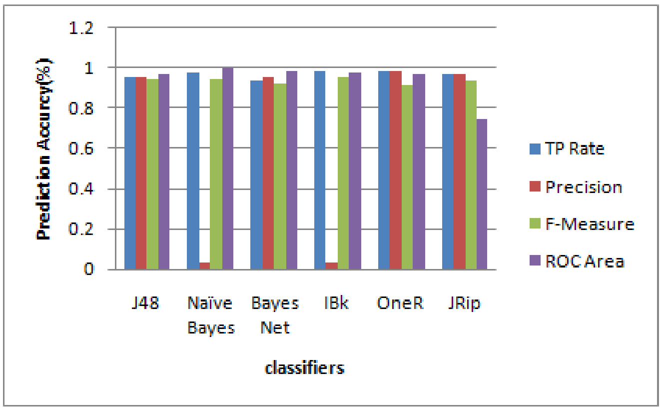


Figure 2: Results of CfsSubset Evaluator

Table 11: Classification results for Chi-Square Attribute Evaluation

| Classifier | TP Rate | Precision | Recall | F-Measure | ROC Area |
|---|---|---|---|---|---|
| J48 | 0.865 | 0.875 | 0.865 | 0.85 | 0.878 |
| Naïve Bayes | 0.955 | 0.958 | 0.955 | 0.956 | 0.986 |
| Bayes Net | 0.921 | 0.931 | 0.921 | 0.925 | 0.98 |
| IBk | 0.955 | 0.959 | 0.955 | 0.956 | 0.953 |
| OneR | 0.983 | 0.984 | 0.983 | 0.983 | 0.967 |
| JRip | 0.972 | 0.973 | 0.972 | 0.971 | 0.954 |

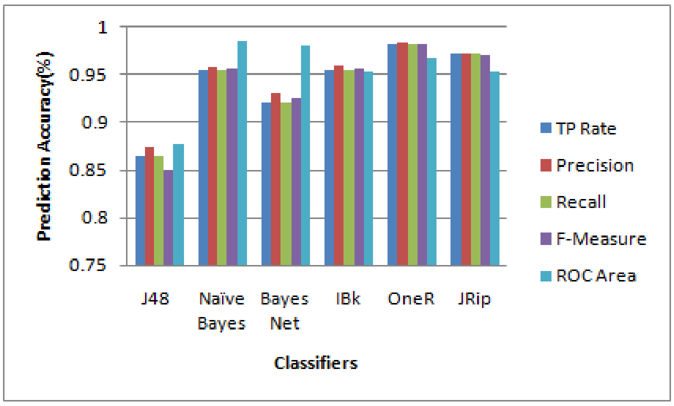


Figure 3: Results of Chi-Square Attribute Evaluation

The graph shows that the classifier Naïve Bayes could attain a highest ROC value and Bayes Net had second highest value when it had 12 features. So, we deduce that Naïve Bayes has the optimal dimensionality in the student data set. Among the classification Algorithm OneR has the maximum F-measure value.

D. Information Gain Attribute Evaluator and Ranker (10 Attributes)

The present study implements Information gain Attribute evaluator on the data set in the WEKA environment and the results are shown in Table 12. Feature selection algorithm Information Gain Attribute Evaluation found that Naïve Bayes classifier correctly classifies about 95.930% for the 10 fold cross-validation testing. The results from Table 12 also reveal that the True Positive Rate and precision value is high for Naïve Bayes and while it is low for J48.

Table 12: Classification results for InfoGain Attribute Evaluation

| Classifier | TP Rate | Precision | Recall | F-Measure | ROC Area |
|---|---|---|---|---|---|
| J48 | 0.907 | 0.913 | 0.907 | 0.907 | 0.935 |
| Naïve Bayes | 0.959 | 0.96 | 0.956 | 0.959 | 0.988 |
| Bayes Net | 0.93 | 0.949 | 0.93 | 0.936 | 0.972 |
| IBk | 0.953 | 0.953 | 0.953 | 0.952 | 0.972 |
| OneR | 0.952 | 0.952 | 0.952 | 0.951 | 0.945 |
| JRip | 0.947 | 0.947 | 0.947 | 0.926 | 0.945 |

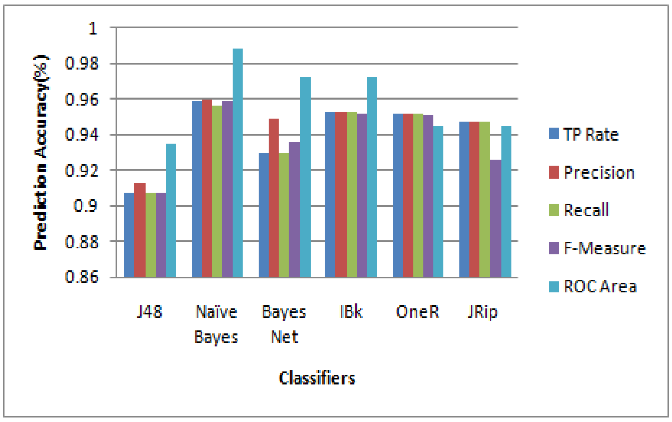


Figure 4: Results of InfoGain Attribute Evaluator

We observer from Fig.5, that classifier Naïve Bayes has highest ROC value and F-measure value found to be maximum in Naïve Bayes with 10 attributes. Therefore, Naïve Bayes can achieve relatively good performance on classification tasks.

E. Relief Attribute Evaluation and Ranker (8 Attributes)

The result from the Table 13 shows that Classifier OneR correctly classifies about 98.181% for the 10-fold Cross-validation testing on the data set and also True positive rate is high. It also presents that Precision and True Positive Rate is low for Bayes Net.

Table 13: Classification results for Relief Attribute Evaluation

| Classifier | TP Rate | Precision | Recall | F-Measure | ROC Area |
|---|---|---|---|---|---|
| J48 | 0.977 | 0.977 | 0.977 | 0.976 | 0.955 |
| Naïve Bayes | 0.958 | 0.962 | 0.958 | 0.958 | 0.987 |
| Bayes Net | 0.915 | 0.938 | 0.915 | 0.923 | 0.969 |
| IBk | 0.97 | 0.972 | 0.97 | 0.97 | 0.98 |
| OneR | 0.982 | 0.982 | 0.982 | 0.981 | 0.966 |
| JRip | 0.964 | 0.965 | 0.964 | 0.962 | 0.963 |

Graphical Representation of Table 13 is shown in fig. 6. We observe that Naïve Bayes could attain the maximum ROC

International Journal of Data Mining Techniques and Applications
Volume: 05 Issue: 02 December 2016 Page No.131-139
ISSN: 2278-2419

value of 0.987 when it had 8 features. And F-measures are highest for the classifier OneR.

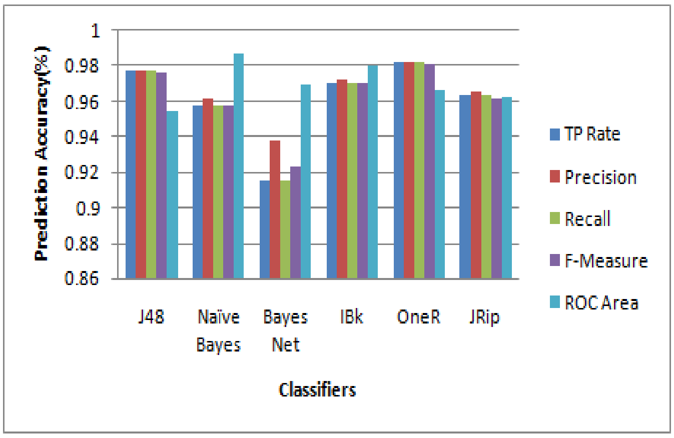


Figure 5: Results of Relief Attribute Evaluation

The results of the comparative study of six different classifiers carried out against feature subsets generated by the four different Feature selection Algorithms are shown below in the form of Table and Graph. In Table 14, we observe that IBK comes up with quite good rules for characterizing the structure in data set. IBK has a highest Accuracy (99.680) over a CfsSubsetEval(CSER). Also Naïve Bayes shows a second highest Accuracy (98.569) among the six Classifiers. Furthermore CfsSubsetEval has the highest accuracy for all the six classifiers than the Chi-Square, InfoGain and Relief Attribute Evaluator.

Table 14: Accuracy of Classifiers over Feature Selection Algorithms

| Classifier | WFS | CSER | CSAER | IGAER | RAER |
|---|---|---|---|---|---|
| J48 | 94.974 | 96.515 | 86.516 | 90.697 | 90.674 |
| Naïve Bayes | 97.989 | 98.569 | 95.505 | 95.930 | 95.757 |
| Bayes Net | 93.969 | 97.670 | 92.134 | 93.023 | 91.515 |
| IBk | 98.492 | 99.680 | 95.405 | 95.255 | 96.969 |
| OneR | 98.442 | 98.112 | 98.314 | 95.254 | 98.181 |
| JRip | 96.984 | 98.478 | 97.191 | 94.674 | 96.363 |

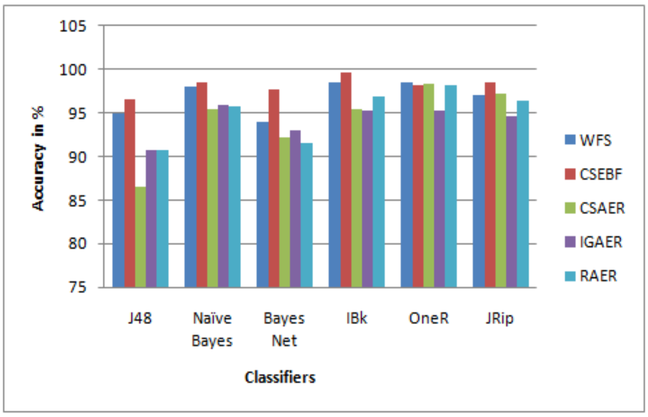


Figure 6: Performance of classifiers over a Feature Selection Algorithms

## VIII. CONCLUSION

The research work aims at analysing the impact of feature selection techniques on the classification task using feature selection algorithms WFS, CSER, CSAER, IGAER and RAER and implementing on the student's dataset collected from three arts and science colleges. The analysis done on the resultant reduced data sets yields faster than models built with no feature selection. The main concentration of this research work is to classify the student's performance in the end semester examination based on their results and personal characteristics. In this paper, it is applied various rank based feature selection filters to data sets to identify the best feature selection algorithm. Based on the results obtained, the performance of feature selection algorithms CFS Subset Evaluator was found to be better than the performance of other three feature selection algorithms. Also, among the classification algorithms, the work identify that IBK algorithm yields 99.68% which better than Naive Bayes, BayesNet, OneR and J48 algorithms. The future work will concentrate on the implementation of a proposed hybrid method by considering large dataset collected from many institutions.

## References

[1] Parneet Kaur,Manpreet Singh,Gurpreet Singh Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector", 3rd Int. Conf. on Recent Trends in Computing, Vol 57, 2015, pp. 500-508.

[2] HythemHashim, Ahmed A, Talab, Ali Satty, Samani A, Talab, "Data Mining Methodologies To Study Students Academic Performance Using the C4.5 Algorithm", Int. Journal on Computational Sciences & Applications, Vol.5, No.2, 2015, pp. 59-68.

[3] Vaibhav P. Vasani, Rajendra D. Gawali, "Classification and performance evaluation using data mining algorithm", Int. Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 3, 2014, pp. 10453-10458.

[4] AnjanaPradeep and Jeena Thomas, "Predicting College Students Dropout using EDM Techniques", Int. Journal of Computer Applications, Vol.123,No.5,2015, pp. 26-34.

[5] PoojaThakar, Anil Mehta, Manisha,"Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue", Int. Journal of Computer Applications, Vol. 110, No.15, Jan 2015, pp.60-68.

[6] Punlumjeak W, Rachburee N, "A Comparative study of feature selection techniques for classify student performance", 7th Int. Conf. on Information Technology and Electrical Engineering, 2015, pp.425-429.

[7] Anal Acharya, Devadatta Sinha, "Application of Feature Selection Methods in Educational Data Mining", Int. Journal of Computer Applications, Vol.103, No.2,2014,pp.34-38.

[8] Komal S,Sahedani, B supriya Reddy, "A Review: Mining Educational Data to Forecast Failure of Engineering Students", Int. Journal of Advanced Research in Computer Science and Software Engineering, Vol.3, Issue 12, 2013, pp. 628-635.

[9] Tarik A. Rashid," Improving on Classification Models of Multiple Classes through Effectual Processes", Int.

Volume: 05 Issue: 02 December 2016 Page No.131-139
ISSN: 2278-2419

Journal of Advanced Computer Science and Applications, Vol.6, No.7, 2015, pp.55-62.

[10] Sadaf Fatima salim Attar, Y.C. Kulkarni, " Precognition of Students Academic Failure Using Data Mining Techniques", Int. Journal of Advanced Research in Computer and Communication Engineering, Vol.4, Issue 5, 2015, pp. 109-115.

[11] Carlos Marquez-Vera, Cristobal Romero Morales, and Sebastian Ventura Soto, "Predicting school failure and dropout by using data mining techniques", IEEE Journal of Latin-American Learning Technologies, Vol. 8, No. 1,2013, pp. 7-14.

[12] Sahil P. Karkhanis, Shweta S. Dumbre, "A Study of Application of Data Mining and Analytics in Education Domain", Int. Journal of Computer Applications, Vol.120, No.22, 2015, pp. 23-28.

[13] Ogunde A.O and Ajibade D.A,"A Data Mining System for Predicting University Students Graduation Grades Using ID3 Decision Tree Algorithm", Journal of Computer Science and Information Technology, Vol. 2, Issue 1, 2014, pp. 21-46.

[14] Dinesh Kumar.A,Radhika.V, "A Survey on Predicting Student Performance", Int. Journal of Computer Science and Information Technologies, Vol.5, No.5,2014,pp.6147-6149.

[15] ArpitTrivedi, "Evaluation of Student Classification Based On Decision Tree", Int.Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 2, 2014, pp. 111-112.

[16] AgrawalBhawana D, GuravBharti B, "Review on Data Mining Techniques Used For Educational System", Int. Journal of Emerging Technology and Advanced Engineering, Vol. 4, Issue 11, 2014, pp. 325-329.

[17] Cao, Longbing, Data mining and multi-agent integration, Springer Science & Business Media, 2009.

[18] Tan, Pang-Ning, Michael Steinbach, and VipinKumar, "Introduction to data mining", Boston: Pearson Addison Wesley, Vol. 1, 2006.

[19] Koller, Daphne, and MehranSahami, "Toward optimal feature selection", 1996.

[20] Guyon, Isabelle, and André Elisseeff, "An introduction to feature extraction", Feature extraction, Springer Berlin Heidelberg, 2006, pp.1-25.

[21] Ramaswami. M and Bhaskaran. R, "A Study on Feature Selection Techniques in Educational Data Mining", Journal of Computing, Vol.1, Issue 1, 2009,pp.7-11.

[22] Forman, George, "An extensive empirical study of feature selection metrics for text classification",The Journalof machine learning research, Vol.3, 2003, pp. 1289-1305.

[23] Dy, Jennifer G., et al., "Unsupervised feature selection applied to content-based retrieval of lung image", Pattern Analysis and Machine Intelligence, IEEE Transactions on Vol.25, No.3, 2003, pp. 373-378.

[24] Egozi, Ofer, EvgeniyGabrilovich, and Shaul Markovitch, "Concept-Based Feature Generation and Selection for Information Retrieval", AAAI, Vol. 8, 2008, pp. 1132-1137.

[25] Yu, Lei, and Huan Liu, "Redundancy based feature selection for microarray data", Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2004, pp. 737-742.

[26] Bolon-Canedo, Veronica, Noelia Sanchez-Marono, and Amparo Alonso-Betanzos, "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset", Expert Systems with Applications, Vol. 38, No.5, 2011, pp. 5947-5957.

[27] Lee, Wenke, Salvatore J. Stolfo, and Kui W. Mok, "Adaptive intrusion detection: A data mining approach", Artificial Intelligence Review, Vol. 14, No.6, 2000, pp. 533-567.

[28] Saari, Pasi, TuomasEerola, and Olivier Lartillot, "Generalizability and simplicity as criteria in feature selection: Application to mood classification in music", Audio, Speech, and Language Processing, IEEE Transactions, Vol. 19, No. 6, 2011, pp.1802-1812.

[29] Mark A. Hall, "Correlation-based Feature Selection for Machine Learning", PhD Thesis, The University of Waikato, 1999.

[30] Mehdi Naseriparsa, Amir-MasoudBidgoli and Tourajvaraee, "A Hybrid Feature Selection Method to Improve Performance of a Group of Classification Algorithms", Int. Journal of Computer Applications, Vol.69, No. 17, pp. 28-35.

[31] Russel, Stuart, and Peter Norvig, "Artificial Intelligence: A Modern Approach ",EUA: Prentice Hall, 2003.

[32] Quinlan, J. Ross,C4. 5: Programs for machine learning, Elsevier, 2014.

[33] Cohen, William W, "Fast effective rule induction", Proceedings of the twelth international conference on machine learning, 1995, pp.115-123.