A Study of Feature Selection Algorithms for Predicting Students Academic Performance

Maryam Zaffar
Department of Computer and Information Sciences
Universiti Teknologi PETRONAS,
32610 Seri Iskander, Malaysia

K.S. Savita
High Performance Cloud Computing Center
Department of Computer and Information Sciences
Universiti Teknologi PETRONAS,
32610 Seri Iskander, Malaysia

Manzoor Ahmed Hashmani
High Performance Cloud Computing Center
Department of Computer and Information Sciences
Universiti Teknologi PETRONAS,
32610 Seri Iskander, Malaysia

Syed Sajjad Hussain Rizvi
Department of Telecommunication
Hamdard University, Karachi,
Pakistan

***Abstract*—The main aim of all the educational organizations is to improve the quality of education and elevate the academic performance of students. Educational Data Mining (EDM) is a growing research field which helps academic institutions to improve the performance of their students. The academic institutions are most often judged by the grades achieved by the students in examination. EDM offers different practices to predict the academic performance of students. In EDM, Feature Selection (FS) plays a vital role in improving the quality of prediction models for educational datasets. FS algorithms eliminate unrelated data from the educational repositories and hence increase the performance of classifier accuracy used in different EDM practices to support decision making for educational settings. The good quality of educational dataset can produce better results and hence the decisions based on such quality dataset can increase the quality of education by predicting the performance of students. In the light of this mentioned fact, it is necessary to choose a feature selection algorithm carefully. This paper presents an analysis of the performance of filter feature selection algorithms and classification algorithms on two different student datasets. The results obtained from different FS algorithms and classifiers on two student datasets with different number of features will also help researchers to find the best combinations of filter feature selection algorithms and classifiers. It is very necessary to put light on the relevancy of feature selection for student performance prediction, as the constructive educational strategies can be derived through the relevant set of features. The results of our study depict that there is a 10% difference of prediction accuracies between the results of datasets with different number of features.**

Keywords*—Educational data mining; feature selection algorithms; classifiers; CFS; relief feature selection algorithm

I. Introduction

Education is a prime factor for the development of a nation. The quality of education is one of the most needed ingredients in creating remarkable members of society. The data kept in academic institution databases plays noteworthy role for the improvement of educational process by exploring the hidden information [1]. Many techniques are being used to evaluate the performance of students' academics. Data Mining techniques are being broadly used on student data these days [2], [3] and is playing a positive role in the area of Educational Data Mining (EDM). EDM discovers the educational data to comprehend the issues in student's academic performance using the fundamental nature of data mining techniques [4]. Student performance prediction is considered as an important topic in EDM. As the performance of student not only effect the organization reputation, but also the future of the student itself, therefore the student performance prediction models are in spot light in front of educational stakeholders. EDM deploys data to help academic organizations in planning educational strategies and in turn enhancing the quality of education.

Student academic progress can be monitored through the prediction models. These prediction models use different EDM techniques to analyze the students' academic performance. It is very hard to distinguish the features affecting the student academic performance [5]. Student academic performance prediction can be helpful for institutions to identify students in need of financial assistance [6], [7], improve institution enrolment quality [7], [8], help students to plan better for future, and also to overcome their struggle with studies. The students' performance prediction model depends on the selected features from the dataset. The most suitable features can be selected by applying feature selection algorithm [9]. These algorithms can refine the prediction results [10]. However, the Feature selection algorithms are best to extract the relevant features and avoid redundancy, without cost of data loss [11], therefore it is very suitable to use FS algorithms in EDM to avoid loss of important data to build strategies with the help of such a quality data.

Feature Selection algorithms are used in in pre-processing step of data. It supports to select the appropriate subset of features to construct a model for data mining. However, Feature Selection algorithms are utilized to improve the predictive accuracy and lower the computational complexity [4], [12], [13]. The feature selection algorithms can increase the performance of student performance prediction models. There are three main types of feature selection algorithms three main categories: filter, wrapper, and hybrid models. Filter method is performed on pre-processing step, and are not depended on any learning algorithm, but they depend on over-all features of the training data. Wrapper method uses learning algorithms to estimate the features. Whereas Hybrid Feature selectin combines the properties of both filter and wrapper method [12]. In this study we focus mainly on the filter feature selection algorithm.

Feature selection has been used in EDM in different research works [5], [9], [14]. Researchers in EDM use different feature selection algorithms to yield effective results in predicting academic performance of students. But still a lot of attention is required to construct student performance prediction models with the help of feature selection algorithms. Our paper is a step towards detecting the best amalgamations of feature section algorithms and classification algorithms on student datasets.

The outline of the paper is as follows: Section II provides the literature related to the feature selection algorithm used in the field of EDM. Section III provides the research methodology followed by the paper. Section IV illustrates the results and discussions. Conclusion of the study is described in Section V.

II. Related Literature

This section gives a brief literature review on the feature selection algorithms used in the field of EDM and the different combinations of feature selection along with classification algorithms used in the other studies. The study in [15] proposed an improved decision tree to predict the indicators of student dropouts. The study collects the dataset of 240 students through a survey and applies Correlation based Feature Selection (CFS) algorithm (Filter feature selection algorithm) in pre-processing step. The classification accuracy of the model shows more than 90%. However, the study took only one dataset into consideration. The investigation in [4] evaluated six feature selection algorithms to predict the performance of higher secondary students. The results of the study conclude that Voted Perceptron, and One Rule (OneR) shows high predictive performance with all the feature subsets gained through feature selection algorithms. Furthermore, Information Gain (IG) and CFS shows better ROC value and F-measure values on higher secondary school dataset.

A study to predict the performance of student in secondary school at Tuzla was presented in [1]. The study used Gain Ratio (GR) feature selection algorithm on the dataset with 19 features. The results with Random Forest classification (RF) algorithm reveals best results in terms of prediction accuracy.

The investigation in [16] was conducted to predict the enrolment of students in Science, Technology, Engineering and Mathematics (STEM) in higher educational institutions in Kenya. Almost 18 features were collected through a questionnaire. The CART decision tree shows better prediction accuracy results with Chi-Square and IG feature selection algorithms.

A study to predict the grades of student was conducted in [14], Principal Component Analysis (PCA) was performed on the dataset of students enrolled in the computer science bachelor's degree . The study uses PCA to build decision trees from the features extracted through the Moodle Logs, to predict student grades.

A comparison between Greedy, IG-ratio ,Chi-Square and mRMR feature selections, was conducted in the study of [17]. The study collected first year students' record with 15 attributes, from the database of University of Technology, Thailand. The study proposed that Greedy Forward selection can give better prediction accuracy result with artificial neural network (ANN) as compared to Naïve Bayes, decision tree and k-NN.

The existing studies in educational data mining have used different filter feature selection algorithms on student datasets. In this study, we used two different datasets, with different number of features. This study is an extension of our previous work [18].

III. Research Methodology

This research article is an extended version of the paper [18]. One dataset was used in previous study to check the performance of different feature selection algorithms. The foremost objective of this research is to estimate the performance of different FS algorithms along with different classification algorithms using different students' datasets with dissimilar number of features. The comparison between the results of FS algorithms is based on two datasets to provide to new educational data mining for the performance of various feature selection algorithms with different number of features. This study will answer two research questions that are:

RQ1. What are the important feature selection algorithms to predict the academic performance of students (Whether they pass or fail)?

RQ2. What are the best possible combinations of feature selection algorithms and classification algorithms to predict the performance of students (Whether they pass or fail)?

To achieve the research objective and to answer the above-mentioned research questions, two student datasets are taken from valid sources, after which different FS algorithms are applied which was not used earlier on this dataset in the previous studies. As in this paper we try to evaluate different feature selection algorithms to check their performance. Various classification algorithms are applied using different FS algorithms. It is evaluated to check the performance among all the combinations applied on students' dataset. Fig. 1 describes a basic flow of our study. Two student datasets were taken in this study. In the second step feature selection algorithms are applied separately on both datasets, in combination of different classification algorithms. Results of precision and correctly classified instances were compared in the final step.

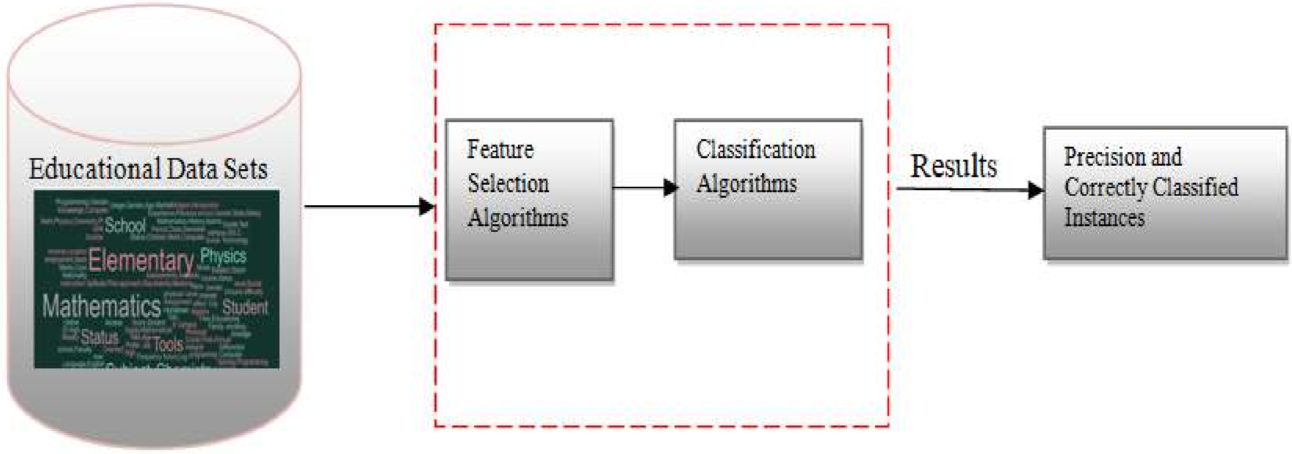

Fig. 1. Flow of methodology.

A. Dataset Description

In this study we have taken two student datasets with different number of features to check the performance of feature selection algorithm on different number of features. The details of two datasets used in this study are given below.

1) Dataset 1: The dataset 1 is comprised of 500 students records with 16 features. This dataset has been used in the study [19], and is available publicly even on Kaggle dataset repository .It is being previously used to check the learner's interactivity with e-learning management system. However, only information gain based feature selection algorithm is used previously. There are three categories of attributes in this dataset demographic, academic and behavioral. The dataset is being used by our previous version of this study.

2) Dataset 2: The dataset 2 is comprised of 300 students with 24 features. It was collected from the three different collages of India. This dataset is used in the study [20] . The dataset is being used in this paper to analyze the student's academic performance.

B. Experimental Setup

Waikato Environment for Knowledge Analysis (WEKA) is developed by University of Waikato in New Zealand as data mining tool. It is built in Java language, and a rich source of data mining algorithms. WEKA offers skill for developing machine learning techniques for different data mining tasks [21], [22]. In this experiment we have used Weka version 3.9, and explorer application.

C. Feature Selection Algorithm and Classifiers

Feature selection is one of the most recurrent and significant technique in data pre-processing and is said to be an essential element of machine learning process [23]. The main focus of our research in this paper is on six important FS algorithm CfsSubsetEval, ChiSquared-AttributeEval, FilteredAttributeEval, GainRatioAttribute-Eval, Principal Components, and ReliefAttributeEval feature selection algorithms are evaluated.

1) CfsSubsetEval: This approach identify the predictive capability of every feature. However, the redundancy factor also plays a critical role in this approach [24], [25]. CFS algorithm uses homogeneous feature in selection process along with discretization preprocessing steps [26].

2) ChiSquaredAttributeEval: Chi-Square is used to compare the tests of independence and the test of goodness of fit. Test of independence estimates whether the class label is dependent or independent of a feature. ChiSquared-AttributeEval estimates an attribute by calculating the value of the chi-squared statistic relating to the class [17], [25].

3) FilteredAttributeEval: This filter feature selection algorithm is available in Weka plate form.

4) GainRatioAttributeEval: The Gain Ratio is the non-symmetrical measure that is introduced to compensate for the bias of the information gain [27]. It is a filter feature selection algorithm that measures how common a feature in a class associated to all other classes.

5) Principal Components: Principal Component analysis reduces the dimensionality of space, without reducing the number of features [28].

6) ReliefAttributeEval: Relief is a simple weight-based algorithm which depends totally on a statistical method. It evaluates the significance of an attribute by sampling an instance repeatedly [25]. It detects those features which are statistically related to the target concept. It has a limitation of non-optimal feature set size [29].

Prediction accuracy of the features selected from the feature selection algorithms can be evaluated through classification algorithms. In our previous work we have used fifteen classification algorithms that are: Bayesian Network (BN), Naïve Bayes (NB), NaiveBayesUpdateable (NBU), MLP, Simple Logistic (SL), SMO, Decision Table (DT), OneR J rip, Decsion Stump (DS), J48, Random Forest (RF), RandomTree (RT), REPtree (RepT). However due the limitation of space we have selected six classification algorithms in this paper.

IV. Results and Discussions

This research reported focuses on the performance evaluation of six Feature Selection algorithms using two different student's datasets. The effectiveness of these algorithms is measured through Precision, Recall, F-measure and prediction accuracy (Correctly classified instances). F-measure is defined as the harmonic mean of precision and recall [30]. The results presented in our previous study [18] and which is then compared with the results obtained usingdataset1 and dataset 2. The outcomes of the six Feature Selection techniques using dataset 1 are reported in Tables I to VI by applying 15 classifiers. These tables illustrate results obtained by each of the Feature Selection (FS) algorithms. Furthermore, each table of results contains four columns that are FS-Classification Algorithm, Precision, Recall and F-measure values.

A. Results on Dataset 1

The results in Table I shows the different values of accuracy measures for fifteen classifiers with Cfssubseteval feature selection algorithm using dataset 1. Fig. 2 graphically illustrates the results obtained with ChiSquared-AttributeEval feature selection algorithms. The results presented in Table II and Fig. 3 depicts that the classifier Decision Stump (DS) has the lowest performance on educational dataset 1 with ChiSquaredAttributeEval; however, MLP classifier shows comparatively better results than other classifiers with the same FS technique.

The results presented in Table III and Fig. 4 indicates that the accuracy of classifiers used on educational data with FilteredAttributeEval feature selection algorithm. The results demonstrate that the values of Precision, Recall and F-measure are comparatively low when Decision Stump and Jip classifiers are applied. While MLP performance is relatively improved than other classifiers using FilteredAttributeEval.

TABLE I. Results of CfsSubsetEval on Dataset 1 using Different Classifiers [18]

FS-Classification Algorithm	**Precision**	**Recall**	**F-Measure**
Cfs-BN	0.724	0.743	0.742
Cfs-NB	0.73	0.729	0.728
Cfs-NBU	0.73	0.729	0.729
Cfs-MLP	0.736	0.729	0.729
Cfs-SL	0.724	0.722	0.723
Cfs-SMO	0.668	0.667	0.667
Cfs-DT	0.693	0.688	0.688
Cfs-Jrip	0.659	0.66	0.658
Cfs-OneR	0.611	0.583	0.571
Cfs-PART	0.713	0.708	0.71
Cfs-DS	0.373	0.528	0.437
Cfs-J48	0.708	0.701	0.702
Cfs-RF	0.64	0.632	0.633
Cfs-RT	0.627	0.618	0.621
Cfs-RepT	0.667	0.66	0.655

TABLE II. Results of ChiSquaredAttributeEval on Dataset 1using Different Classifiers [18]

FS-Classification Algorithm	**Precision**	**Recall**	**F-Measure**
Chi-BN	0.716	0.715	0.716
Chi-NB	0.66	0.66	0.654
Chi-NBU	0.66	0.66	0.654
Chi-MLP	0.769	0.764	0.764
Chi-SL	0.715	0.708	0.709
Chi-SMO	0.741	0.736	0.737
Chi-DT	0.71	0.701	0.702
Chi-Jrip	0.698	0.694	0.692
Chi-OneR	0.611	0.583	0.571
Chi-PART	0.64	0.639	0.639
Chi-DS	0.373	0.528	0.437
Chi-J48	0.709	0.708	0.708
Chi-RF	0.718	0.715	0.716
Chi-RT	0.674	0.674	0.674
Chi-RepT	0.651	0.653	0.651

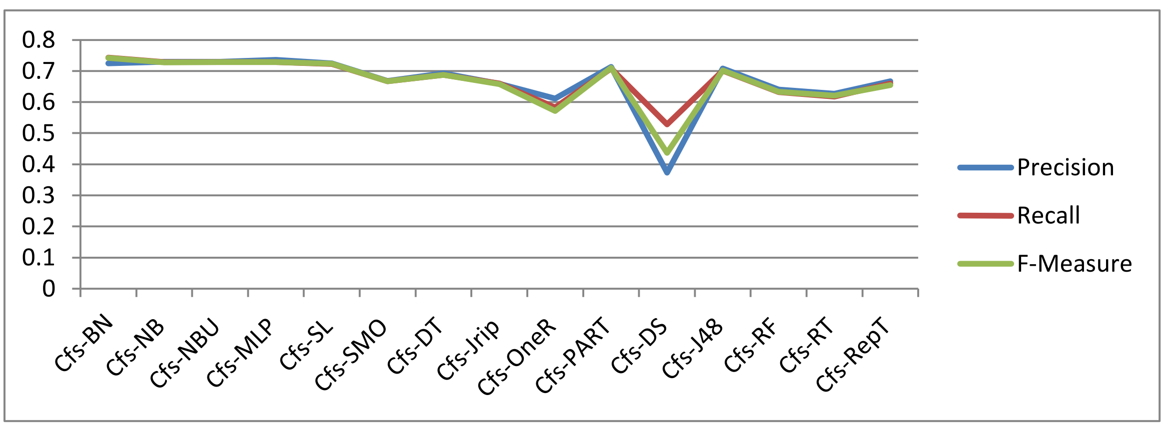

Fig. 2. Performance of CfsSubsetEval using Dataset 1.

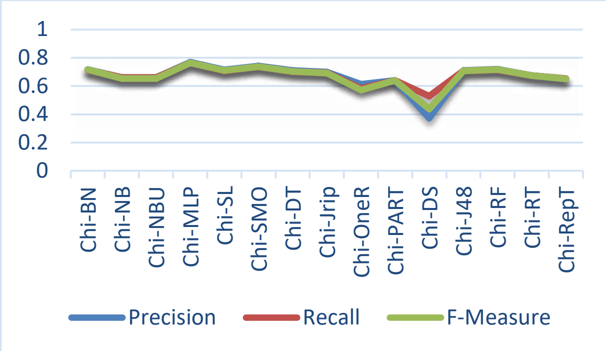

Fig. 3. Performance of ChiSquredAttributeEval using Dataset 1.

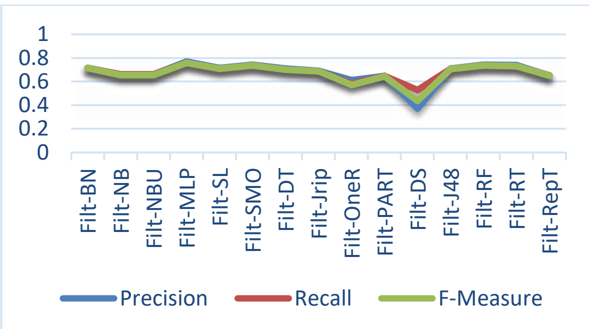

Fig. 4. Performance of FilteredAttributeEval using Dataset 1.

TABLE III. PERFORMANCE EVALUATION OF FILTERED ATTRIBUTE EVAL USING PRECISION RECALL AND F-MEASURE ON DATASET 1 [18]

FS-Classification Algorithm	Precision	Recall	F-Measure
Filt-BN	0.716	0.715	0.716
Filt-NB	0.66	0.66	0.654
Filt-NBU	0.66	0.66	0.654
Filt-MLP	0.768	0.757	0.758
Filt-SL	0.715	0.708	0.709
Filt-SMO	0.741	0.736	0.737
Filt-DT	0.71	0.701	0.702
Filt-Jrip	0.691	0.688	0.688
Filt-OneR	0.611	0.583	0.571
Filt-PART	0.646	0.646	0.645
Filt-DS	0.373	0.528	0.437
Filt-J48	0.709	0.708	0.707
Filt-RF	0.741	0.736	0.737
Filt-RT	0.738	0.729	0.73
Filt-RepT	0.651	0.653	0.651

The results reported in Table IV and Fig. 5 are exhibiting the identical performance details as illustrated earlier in Table III and Fig. 4. The results show that the decrease in performance by applying GainRatioAttributeEval Jrip classifier, however, MLP and SMO performed comparatively better than other classifiers.

The results in Table V present the performance of Principal Components using fifteen selected classification algorithms. Fig. 6 is the graphical representation of the performance of Principal Components. The result in Table V depicts that SMO classifier performed relatively better, while the performance of Jrip and Decision Stump classifiers is contradictory to the expected with Principal component.

TABLE IV. PERFORMANCE EVALUATION OF GAINRATIOATTRIBUTEEVAL USING PRECISION RECALL AND F-MEASURE ON DATASET 1 [18]

FS-Classification Algorithm	Precision	Recall	F-Measure
GR-BN	0.716	0.715	0.716
GR-NB	0.66	0.66	0.654
GR-NBU	0.66	0.66	0.654
GR-MLP	0.768	0.757	0.758
GR-SL	0.715	0.708	0.709
GR-SMO	0.741	0.736	0.737
GR-DT	0.71	0.701	0.702
GR-Jrip	0.691	0.688	0.688
GR-OneR	0.611	0.583	0.571
GR-PART	0.646	0.646	0.645
GR-DS	0.373	0.528	0.437
GR-J48	0.709	0.708	0.707
GR-RF	0.741	0.736	0.737
GR-RT	0.738	0.729	0.73
GR-RepT	0.651	0.653	0.651

TABLE V. RESULTS OF PRINCIPALCOMPONENTS ON DATASET 1USING DIFFERENT CLASSIFIERS [18]

FS-Classification Algorithm	Precision	Recall	F-Measure
PC-BN	0.643	0.632	0.633
PC-NB	0.508	0.507	0.506
PC-NBU	0.508	0.507	0.506
PC-MLP	0.694	0.694	0.693
PC-SL	0.692	0.688	0.688
PC-SMO	0.745	0.736	0.737
PC-DT	0.633	0.618	0.617
PC-Jrip	0.57	0.549	0.545
PC-OneR	0.445	0.444	0.445
PC-PART	0.591	0.59	0.591
PC-DS	0.345	0.486	0.403
PC-J48	0.674	0.667	0.668
PC-RF	0.701	0.694	0.695
PC-RT	0.585	0.576	0.576
PC-RepT	0.659	0.66	0.659

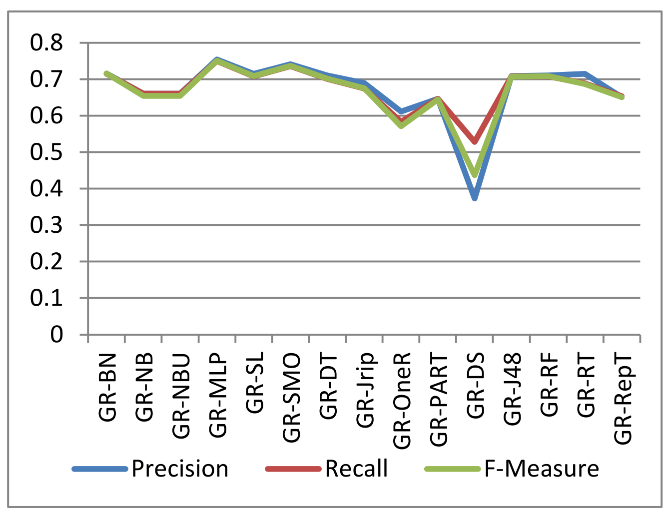

Fig. 5. Performance of GainRatioAttributeEval using dataset 1.

Table VI and Fig. 7 presents the result of ReliefAttributeEval(Rel) using different classifiers. It is observed through the results analysis that Random Forest classifiers shows better results with ReliefAttributeEval, however, the Decision Stump (DS) classifier depicts poor performance with ReliefAttributeEval using data set1 of students records.

B. Comparison of Results on Dataset 1 and Dataset 2

The comparison between the correctly classified instances using dataset 1 and dataset 2 are illustrated in Table VII. In this table six classifiers are presented only which performed better as compared to the other classifiers. The results indicate significant difference in the performance using both the datasets. There is approximately 10 to 20% performance and accuracy difference with each of the FS algorithm.

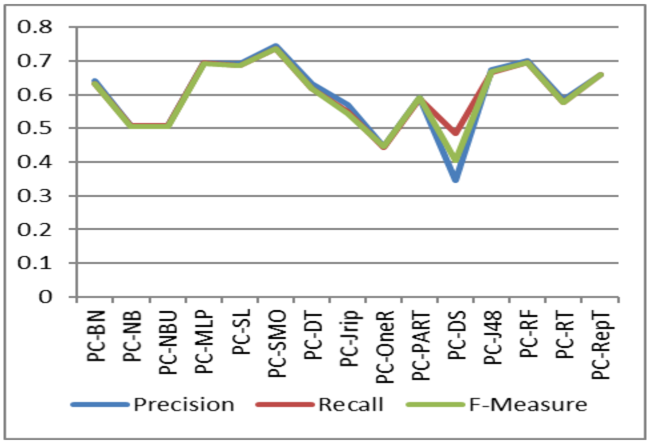

Fig. 6. Precision, recall and F-measure of principal components.

TABLE VI. RESULTS OF RELIEF ATTRIBUTE ON DATASET 1 USING DIFFERENT CLASSIIERS [18]

FS-Classification Algorithm	Precision	Recall	F-Measure
Rel-BN	0.716	0.715	0.716
Rel-NB	0.66	0.66	0.654
Rel-NBU	0.66	0.66	0.654
Rel-MLP	0.767	0.764	0.764
Rel-SL	0.715	0.708	0.709
Rel-SMO	0.741	0.736	0.737
Rel-DT	0.71	0.701	0.702
Rel-Jip	0.713	0.708	0.708
Rel-OneR	0.611	0.583	0.571
Rel-PART	0.646	0.646	0.645
Rel-DS	0.373	0.528	0.437
Rel-J48	0.709	0.708	0.707
Rel-RF	0.756	0.75	0.873
Rel-RT	0.665	0.66	0.657
Rel-RepT	0.651	0.653	0.651

1) Feature Selection Algorithms Accuracy: Relief feature selection and Chi-Square algorithm with MLP classifier provides maximum accuracy using the dataset 1. While dataset 2 is used with chi feature selection technique in combination with Bayes Net (BN) classification algorithm offers the maximum accuracy. Principal component feature reduction technique in combination with Naïve Bayes (NB), provides least accuracy on dataset 1. Though other selected FS techniques in combination with decision tree algorithm exhibits the least accuracy. Hence, the overall performance degrades for dataset 1 with the combination of FS technique and Decision Tree (DT) classifiers. Likewise, the Chi-square FS algorithm with Decision tree results in least performance on the dataset 2. It is concluded from the accuracy measures illustrated in Table VII that performance is better with 16 features of dataset 1 than the 24 features of dataset 2.

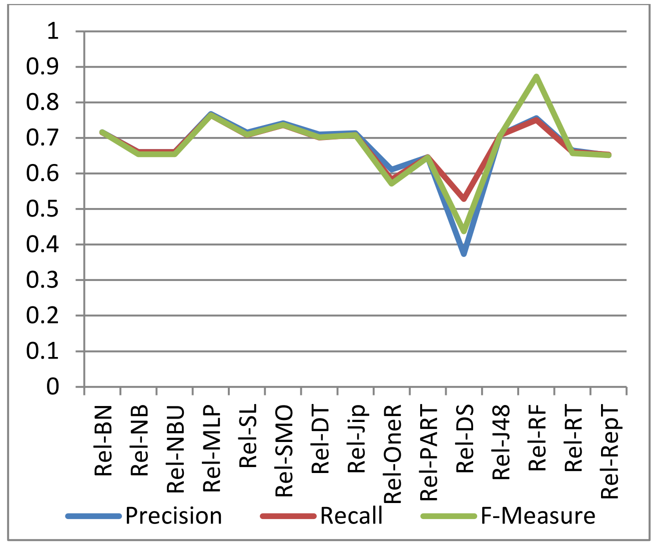

Fig. 7. Performance of ReliefAttributeEval using Dataset1.

A comparative analysis based on the number of features selected in the dataset 1 and dataset 2 with respect to the precision is presented in Table VIII. The chi-square FS technique with Mlp classifiers results in maximum precision using the dataset 1 whereas Cfs algorithm along with the Bayes Net and Naïve Bayes provides maximum precision using the dataset 2. However, the performance of FS techniques with decision tree classification algorithm degrades using the dataset 1 and 2. The performance analysis discussed answer the two research questions discussed in Section III. These results give the answer of two research questions.

RQ1. What are the important feature selection techniques to predict the performance of students?

It is concluded from Tables VII and VIII that the performance of FS techniques has been improved using dataset 1 as compared to the dataset 2. Relief feature selection technique and Chi square algorithm perform better on dataset 1. Whereas Chi square and Cfs feature selection techniques perform better on dataset 2. Hence, these techniques must be considered in predicting the performance of students. According to the analysis, relief, chi-square and cfs are important FS techniques to predict the performance of student.

RQ2. What are the best possible combinations of feature selection techniques and Classification algorithms to predict the performance of students?

Fig. 8 and 9 shows that there is an evident difference in the results of dataset 1 and dataset 2. The results with the dataset 1 are much better than the results with the dataset 2. Both the figures are presenting a clear picture of the results.

TABLE VII. PERFORMANCE EVALUATION OF FEATURE SELECTION ALGORITHMS ON DATASET 1 & 2 IN CONTEXT WITH % OF CORRECTLY CLASSIFIED INSTANCES

FS-Classification Technique	Data set1	Dataset2
Cfs-BN	0.724	0.625
Cfs-NB	0.73	0.625
Cfs-MLP	0.736	0.561
Cfs-SMO	0.668	0.523
Cfs-DS	0.373	0.287
Cfs-RF	0.64	0.614
Chi-BN	0.716	0.616
Chi-NB	0.66	0.597
Chi-MLP	0.769	0.441
Chi-SMO	0.741	0.548
Chi-DS	0.373	0.367
Chi-RF	0.718	0.452
Filt-BN	0.716	0.61
Filt-NB	0.66	0.614
Filt-MLP	0.768	0.496
Filt-SMO	0.741	0.534
Filt-DS	0.373	0.287
Filt-RF	0.741	0.438
GR-BN	0.716	0.559
GR-NB	0.66	0.555
GR-MLP	0.754	0.506
GR-SMO	0.741	0.519
GR-DS	0.373	0.287
GR-RF	0.71	0.565
PC-BN	0.643	0.367
PC-NB	0.508	0.488
PC-MLP	0.694	0.436
PC-SMO	0.745	0.495
PC-DS	0.345	0.28
PC-RF	0.701	0.363
Rel-BN	0.716	0.58
Rel-NB	0.66	0.596
Rel-MLP	0.767	0.439
Rel-SMO	0.741	0.444
Rel-DS	0.373	0.287
Rel-RF	0.756	0.499

TABLE VIII. PERFORMANCE EVALUATION OF FEATURE SELECTION ALGORITHMS ON DATASET 1 & 2 IN CONTEXT WITH % OF CORRECTLY CLASSIFIED INSTANCES

FS-Classification Technique	Dataset1	Dataset2
Cfs-BN	74.31	57.84
Cfs-NB	72.08	55.88
Cfs-MLP	72.92	57.84
Cfs-SMO	66.67	55.88
Cfs-DS	52.78	42.51
Cfs-RF	63.19	59.8
Chi-BN	71.52	61.33
Chi-NB	65.97	59.33
Chi-MLP	76.39	44.33
Chi-SMO	73.61	55
Chi-DS	52.78	42
Chi-RF	71.53	45.33
Filt-BN	71.53	59.8
Filt-NB	65.97	59.8
Filt-MLP	75.69	48.03
Filt-SMO	73.61	51.96
Filt-DS	52.78	42.15
Filt-RF	73.61	42.15
GR-BN	71.53	56.33
GR-NB	65.97	55.66
GR-MLP	75	51
GR-SMO	65.97	54.3
GR-DS	52.78	42.15
GR-RF	70.83	55.88
PC-BN	63.19	45.09
PC-NB	50.69	51.96
PC-MLP	69.44	45.09
PC-SMO	73.61	49.01
PC-DS	48.61	43.13
PC-RF	69.44	47.05
Rel-BN	71.53	55.88
Rel-NB	65.97	53.92
Rel-MLP	76.39	46.07
Rel-SMO	73.61	48.03
Rel-DS	52.78	42.15

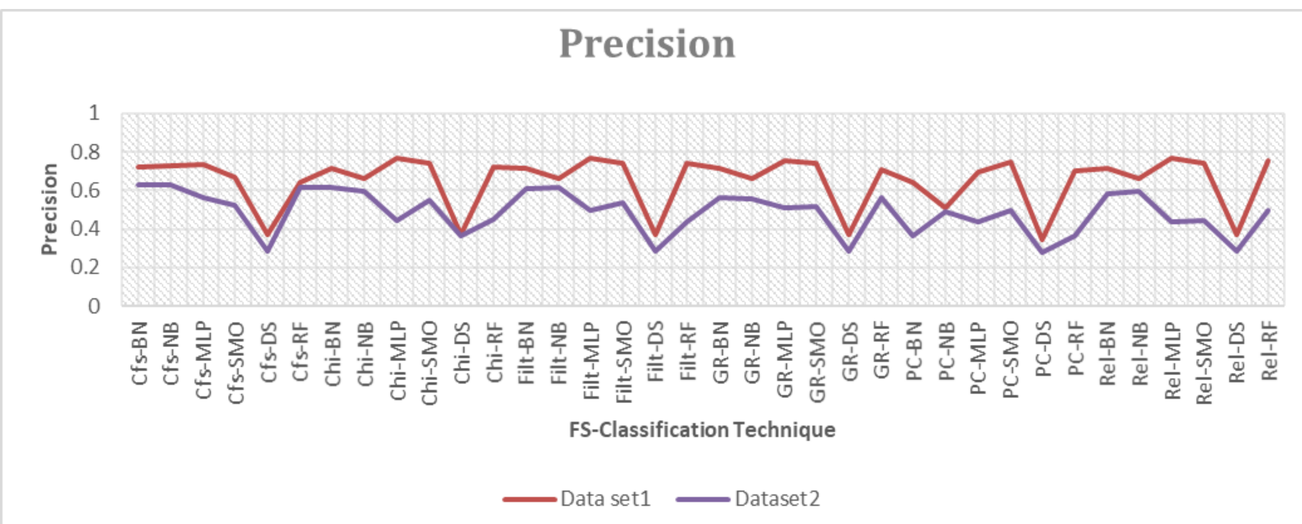

Fig. 8. Comparison of precision accuracy using dataset 1 & 2.

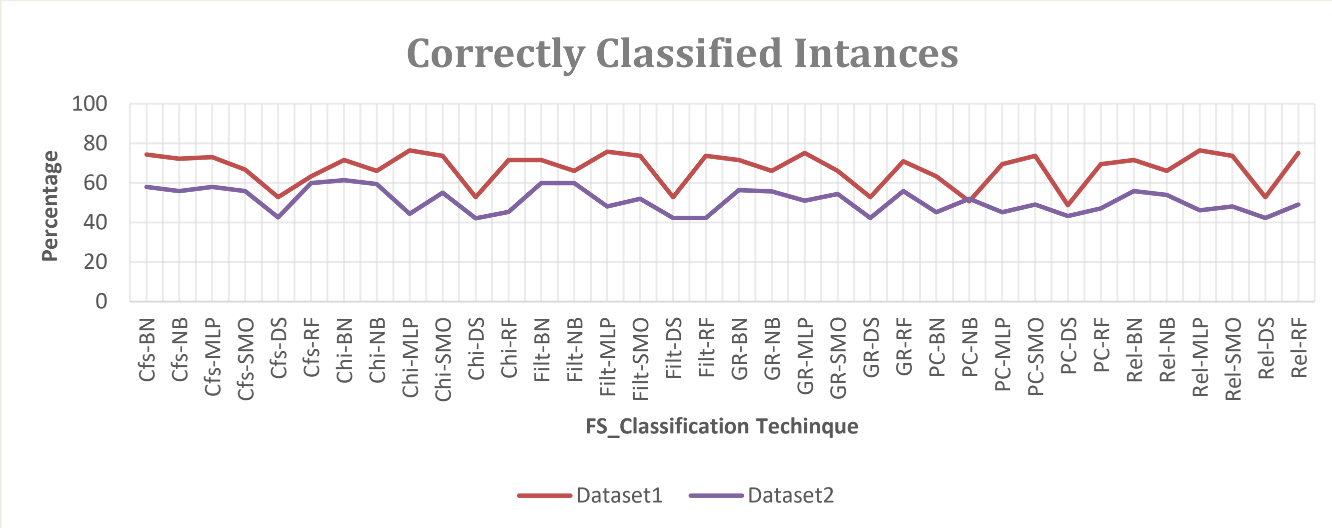

Fig. 9. Comparison of correctly classified instances using dataset 1 & 2.

V. Conclusion

This paper presents the study of various feature selection algorithms and analysed their performance using two different datasets. The results indicated that there is significant performance difference of feature selection algorithms using the datasets with different numbers of features; shows 10 to 20 per cent difference in accuracy percentages. The performance of the filter feature selection techniques reduces as the number of feature increases. To predict the academic performance of the student, having a large number of feature sets, wrappers feature selection techniques can also be evaluated. In future we will also evaluate the feature selection results through confusion m. Furthermore, we cannot neglect the advantages of filter feature selection techniques. In future, the study can be enhanced by applying few hybrid feature selection algorithms on student datasets in order to predict the performance of the student.

References

[1] E. Osmanbegović, M. Suljić, and H. Agić, "DETERMINING DOMINANT FACTOR FOR STUDENTS PERFORMANCE PREDICTION BY USING DATA MINING CLASSIFICATION ALGORITHMS," Tranzicija, vol. 16, pp. 147-158, 2015.

[2] A. M. Shahiri and W. Husain, "A review on predicting student's performance using data mining techniques," Procedia Computer Science, vol. 72, pp. 414-422, 2015.

[3] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, pp. 601-618, 2010.

[4] M. Ramaswami and R. Bhaskaran, "A study on feature selection techniques in educational data mining," arXiv preprint arXiv:0912.3924, 2009.

[5] A. Mueen, B. Zafar, and U. Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques," International Journal of Modern Education and Computer Science, vol. 8, p. 36, 2016.

[6] M. Ramaswami and R. Rathinasabapathy, "Student Performance Prediction," International Journal of Computational Intelligence and Informatics, vol. 1, 2012.

[7] N. T. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of techniques for predicting academic performance," in Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual, 2007, pp. T2G-7-T2G-12.

[8] P. Golding and O. Donaldson, "Predicting academic performance," in Frontiers in education conference, 36th Annual, 2006, pp. 21-26.

[9] H. M. Harb and M. A. Moustafa, "Selecting optimal subset of features for student performance model," Int J Comput Sci, p. 5, 2012.

[10] M. Doshi, "Correlation Based Feature Selection (Cfs) Technique To Predict Student Perfromance," International Journal of Computer Networks & Communications, vol. 6, p. 197, 2014.

[11] W. Punlumjeak and N. Rachburee, "A comparative study of feature selection techniques for classify student performance," in Information Technology and Electrical Engineering (ICITEE), 2015 7th International Conference on, 2015, pp. 425-429.

[12] D. Koller and M. Sahami, "Toward optimal feature selection," Stanford InfoLab1996.

[13] P. Mitra, C. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," IEEE transactions on pattern analysis and machine intelligence, vol. 24, pp. 301-312, 2002.

[14] A. Figueira, "Predicting Grades by Principal Component Analysis: A Data Mining Approach to Learning Analyics," in Advanced Learning Technologies (ICALT), 2016 IEEE 16th International Conference on, 2016, pp. 465-467.

[15] S. Sivakumar, S. Venkataraman, and R. Selvaraj, "Predictive Modeling of Student Dropout Indicators in Educational Data Mining using Improved Decision Tree," Indian Journal of Science and Technology, vol. 9, 2016.

[16] K. W. Stephen, "Data Mining Model for Predicting Student Enrolment in STEM Courses in Higher Education Institutions," 2016.

[17] N. Rachburee and W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," in Information Technology and Electrical Engineering (ICITEE), 2015 7th International Conference on, 2015, pp. 420-424.

[18] M. Zaffar, M. A. Hashmani, and K. Savita, "Performance analysis of feature selection algorithm for educational data mining," in Big Data and Analytics (ICBDA), 2017 IEEE Conference on, 2017, pp. 7-12.

[19] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Mining Educational Data to Predict Student's academic Performance using Ensemble Methods," International Journal of Database Theory and Application, vol. 9, pp. 119-136, 2016.

[20] S. Hussain, N. A. Dahan, F. M. Ba-Alwi, and N. RIBATA, "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA," Indonesian Journal of Electrical Engineering and Computer Science, vol. 9, 2018.

[21] K. Patel, J. Vala, and J. Pandya, "Comparison of various classification algorithms on iris datasets using WEKA," Int. J. Adv. Eng. Res. Dev.(IJAERD), vol. 1, 2014.

[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," ACM SIGKDD explorations newsletter, vol. 11, pp. 10-18, 2009.

[23] A. Kalousis, J. Prados, and M. Hilario, "Stability of feature selection algorithms: a study on high-dimensional spaces," Knowledge and information systems, vol. 12, pp. 95-116, 2007.

[24] M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," 1998.

[25] C. Anuradha and T. Velmurugan, "Feature Selection Techniques to Analyse Student Acadamic Performance using Naïve Bayes Classifier," in The 3rd International Conference on Small & Medium Business, 2016, pp. 345-350.

[26] C. Huertas and R. Juárez-Ramírez, "Filter feature selection performance comparison in high-dimensional data: A theoretical and empirical analysis of most popular algorithms," in Information Fusion (FUSION), 2014 17th International Conference on, 2014, pp. 1-8.

[27] J. Novaković, "Toward optimal feature selection using ranking methods and classification algorithms," Yugoslav Journal of Operations Research, vol. 21, 2016.

[28] Q. Guo, W. Wu, D. Massart, C. Boucon, and S. De Jong, "Feature selection in principal component analysis of analytical data," Chemometrics and Intelligent Laboratory Systems, vol. 61, pp. 123-132, 2002.

[29] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in Aaai, 1992, pp. 129-134.

[30] T. Velmurugan and C. Anuradha, "Performance Evaluation of Feature Selection Algorithms in Educational Data Mining," Performance Evaluation, vol. 5, 2016.