International Journal of Recent Research and Review, Vol. XI, Issue 1, March 2018

ISSN 2277 – 8322

Evaluation of Student Performance Prediction Models with Two-Class Using Data Mining Approach

Rahul Misra[1], Dr. Ramkrishan Sahay[2]

[1]PhD Scholar, Department of CSE, Mahamaya Technical University, Noida, India

[2]Professor, Department of CSE, Mahamaya Technical University, Noida, India

misra.rrahul@gmail.com

Abstract- **The academic achievement of higher secondary school education in India is a turning point in the life of any student, as it serves as a very important link between the higher and higher secondary education of students. But, there are determinants like demographic, academic and socio-economic factors of students that restrict the students' performance. In this paper present the evaluation of student performance prediction models with two-class using data mining approach..**

Keywords- **Data Mining, Education, Student, Performance, predication Models.**

I. INTRODUCTION

Education is a process of imparting or acquiring knowledge and habits through instruction or study and this process results in desirable changes in the behavior of human beings. It provides the skills to individuals to become self-confident, self-reliant and self- sustained and inculcates buoyancy to face challenges in all walks of life. It enhances the ability of individuals to manage health problems, improve nutrition and childcare, and prepare for the future. It sustains the human values which contribute to individual and collective well-being. It is the key which allows people to move up in the world, seek better jobs, and ultimately succeed in their lives. It is essential for eradicating poverty and it allows people to be more productive playing greater roles in economic life and earn a better living. It is worth mentioning that education forms the basis for lifelong learning in the context of human development and it is one of the fundamental requirements of democracy. It makes the people to aware of opportunities and rights that in turn result in more responsible and informed citizens. These citizens can have a voice in politics and society, which is essential for sustaining democracy and so education, is the only tool which takes the country to greater heights.

Besides, educators could also monitor their student's achievements. Students could improve their learning activities, allowing the administration to improve the systems performance. Thus, the application of data mining techniques can be focused on specific needs with different entities. In order to encounter the problems, a systematically review is proposed. The proposed systematically review is to support the objectives of this study, which are:

- To study and identify the gaps in existing prediction methods.
- To study and identify the variables used in analyzing students performance.
- To study the existing prediction methods for predicting students performance.

II. STUDENT PERFROMANCE CLASSIFIERS

In classification, one of the main objectives is to assign a student to a pre-defined class with minimum error rate. A simple example of a classification problem is the evaluation of performance of a student. Given the particulars of a student, a system chooses the most appropriate related class to which the student belongs.

At present, the number of various classification algorithms is enormous and these algorithms contain contributions from many research areas. The performance of each varies widely depending on the application domain and the criteria used in assessing them. When constructing a model for a new problem, most of the researchers prefer to select a most commonly used as well as a more familiar one according to their particular experience and preference.

The applicability of algorithms for a particular application domain has always been a central issue in the field of data mining. There are a number of drawbacks. First, a major concern is that the applicability of algorithms for student performance prediction depends on the nature of data set as well as the volume of data. Furthermore, the nature of these algorithms is such that comparison and characterization of algorithms is hard to achieve. We attempt not only to compare various classification algorithms on the same student dataset by varying number of values of class variables but also to measure the important performance metrics of the classifier algorithms.

This research proposes a general framework to recommend an appropriate classifier algorithm for student performance prediction. There are many different classification algorithms that are evolved from different areas in statistics, machine learning and neural networks. Thus, it becomes a difficult task for a data analyst to cope up with the progress and to make it difficult to select the best (or appropriate) algorithm. The problem is made more difficult by the fact that system designers who are not data analysts want to have access to classification algorithms to assist them in their decision making.

Ideally, it would be optimal, if we are able to identify the single best algorithm, which could be used for all situations. A brute force approach to this problem is to try all the classification algorithms on the dataset with different class values and then select the one with the best results. But, in practice, this approach is not viable in most applications. This is due to the fact that there are too many algorithms to try, some of which may be quite slow in terms of time taken for model construction. The problem is exacerbated when dealing with large amounts of data as is common in knowledge discovery. Another problem with this approach is that a person must know how to use all the algorithms by supplying the appropriate parameters.

Classification accuracy is an important metric about the goodness of a classifier, but this is not the only criterion. Many other evaluation metrics can also provide useful information in comparing the performance of classifiers, such as ROC curve and misclassification cost.

III. RESULTS AND DISCUSSION

To start with, we compared the predictive accuracy of the five classification algorithms J48, Decision Table (DT), BayesNet and NaiveBayes and Multi-Layered Back-propagation (MLP) algorithms with different subset of features obtained through various filter-based, wrapper-based, and hybrid-based feature selection algorithms. The performance of the five classifiers has been discussed below individually for each of the feature selection approach with varying class values.

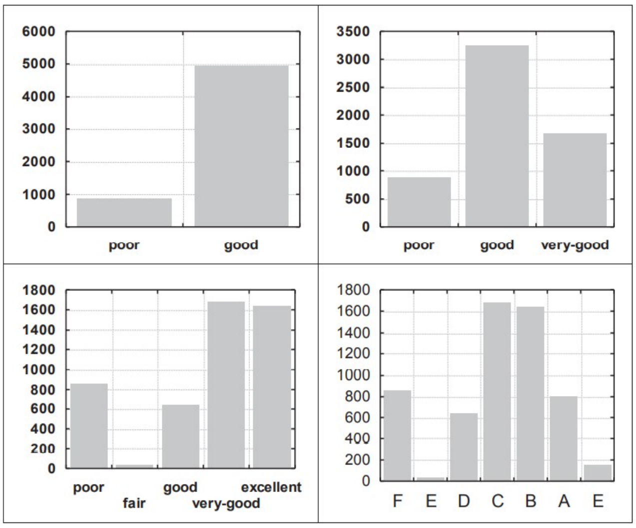

Fig. 1: Histograms for class variable- HSc Grade (2-Level/ 3-Level/ 5-Level and 7-Level Classification)

Evaluation of Two-Class Student Model based on Filtered Feature Subsets

Before going to the evaluation of the performance of student prediction models on their misclassification cost, first we analyzed the merits of the models, based on their predictive accuracy. Table 1 shows the percentage of predictive accuracy of each of the classifier with five filter-based feature selection methods – Full Feature Set (FFS), Correlation-Based. Feature evaluation method (CFS), Consistency Subset evaluation method (CSS), Chi-Square evaluation method (CHI), Gain Ratio Evaluation method (GAR), and Information Gain Evaluation method(ING).

Table 1. Comparing Predictive Accuracy of five classifiers with six filtered feature subsets for two class

Models/ FSS	**FFS**	**CFS**	**CSS**	**CHI**	**GAR**	**ING**
J48	90.0242	85.2261	88.2292	88.0566	87.3317	87.6942
DT	83.2068	85.2951	83.8454	84.1042	83.7936	83.5002
Bayes Net	85.2261	85.2261	85.2261	85.0535	85.0535	85.0535
Naïve Bayes	75.8026	84.3804	80.7732	77.3731	80.2554	77.9945
MLP	87.1246	86.3652	98.5847	97.5837	93.3379	95.7887

The overall results of classifier's performance against with different filter-based feature subset evaluation methods are shown in Fig 2. Regarding individual classifier, for two class student data, MLP had superior performance over other four classifiers. Also the Fig 2 indicates that filter-based feature selection methods highly influenced the predictive accuracy of all the classifiers against Full Feature Set (FSS). The CSS filter method overlooked all other filter methods and provided high predictive accuracy to all classifiers. The features which were influencing the predictive accuracy for two-class student model in CSS filter method were LOC- SCH, BMI, PSEdu, EEdu, XMARK-P, FAM-SIZE, NO-EB, NO-ES, NO-YS, TransSchool, Veh-Home,

LArea, PTution, SpIndoor, SpIOutdoor, FEDU, MEDU and FOCC. The size of the subset produced by CSS was 19, which was higher than the size of the other filtering methods.

In this scheme, MLP-CSS classifier method yielded the highest predictive accuracy of 98.5847%, while J48-CSS yielded 88.2292%. It is also worth mentioning here that the behavior of all the classifiers were at the same level of predictive accuracy as far as CFS filter method was concerned and its accuracy ranged from 84% to 86%. The range of predictive accuracy in CFS method for all the classifiers was lower by 2% to 5% than that of CSS method with just six features. From these observations, we could claim that the filtered feature selection method had enormous impact on the predictive accuracy of student performance model with respect to five classifiers – j48, DT, BayesNet, NaiveBayes and MLP.

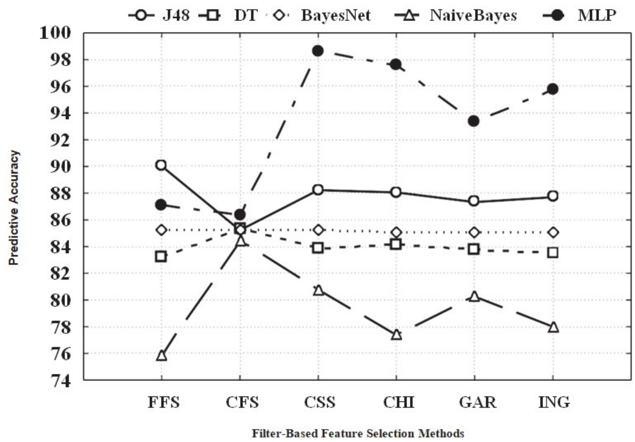

Fig. 2: 2D-line plot showing comparative performance of five classifiers - J48, DT, BayesNet, NaiveBayes and MLP with five filtered feature selection methods for two class student data set

Evaluation of Two-Class Student Model based on Wrapper Feature Subsets

The predictive accuracy of the five classifiers based on the feature subset obtained through Wrapper subset evaluation method is shown in Table 2. The merits of the feature subset were estimated by using NaiveBayes as a base line classifier. Except the Classifier J48, all other classifiers provided high predictive accuracy on the feature subset generated through the Wrapper method against Full Feature Set (FFS). It is pertinent to notice that wrapper- based feature selection method also fine tuned the performance of the classifier against using original data set. The overall predictive accuracy of NB-BF wrapper method was one percent less than that of the filter based CSS method with same subset size. Both NaiveBayes and MLP yielded 10% higher predictive accuracy in wrapper method than that of classifier with FFS.

Table 2. Comparing Predictive Accuracy of five classifiers with NaiveBayes Wrapper feature subsets for two class

Models/FSS	FFS	NB-BF
J48	90.0242	87.7632
DT	83.2068	**84.1733**
BayesNet	85.2261	**85.2261**
NaiveBayes	75.8026	**83.6382**
MLP	87.1246	**97.1004**

Evaluation of Two-Class Student Model based on Hybrid Feature Subsets

Based on the F1-Measure and ROC value, we fixed the number of features in each subset generated through CFS, CHI and ING hybrid methods. Using the results of these subsets, the predictive accuracy has been calculated and it is shown in Table 3. The main purpose of the rank-based method was to achieve the higher predictive accuracy with minimum cardinality of the feature subset. The highlighted predictive measures indicated that better results were obtained with minimum number of features.

Table 3. Comparing Predictive Accuracy of five classifiers with different ranked feature subsets for two class

Models/ FSS	FFS	F1M-CHI-10	F1M-CFS-17	F1M-ING-13	ROC-CHI-9	ROC-CFS-9	ROC-ING-9
J48	90.0242	86.7449	88.4881	87.1419	86.7967	86.486	85.2261
DT	83.2068	83.9144	83.6037	83.552	83.7763	83.5347	85.2951
Bayes Net	85.2261	85.0535	85.0535	85.0535	85.0535	85.1053	85.2261
Naïve Bayes	75.8026	81.9296	79.8412	80.1174	82.3438	83.621	83.4311
MLP	87.1246	91.1288	94.7877	94.2354	89.9206	88.1602	87.1591

The performance of the five different classifiers carried out against feature subsets generated by the six different feature selection procedures are shown in Fig 3. Clearly MLP performed well for the feature subset F1M-CFS-17 generated by Correlation-Based method.

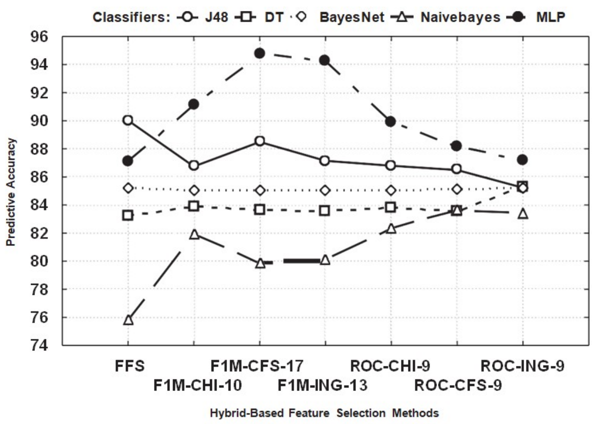

Fig. 3: 2D2D-line plot showing comparative performance of five classifiers- J48, DT, BayesNet, NaiveBayes and MLP with five six hybrid-based feature selection methods for Two-Class Student data set

Also the classifiers NaiveBayes and DT attained high predictive accuracy in ROC-CFS-9 and ROC-ING-9 methods respectively. The predictive value of DT model was higher than the value obtained in both Filter-based and Wrapper-based method with just top ranked features generated by Information Gain evolution method. However, these hybrid based feature selection method did not show any influence on the predictive measure of BayesNet and J48 classifiers. Thus from these investigation, we observed that the three classifiers – MLP, NaiveBayes and DT, with minimal cardinality of feature subset for two-class student data set had significantly higher predictive accuracy.

Since predictive accuracy we not a better evaluation measure for un-balanced data set, we introduced misclassification cost measure mentioned in next section not only to overcome this uncertainty but also to assess the classifier efficiency on student data set. The critically ranking value of zero was fixed for actual class (for "O" class) and it is shown in Table 4.

Table 4. Relative Grade Ranking for Seven-Class Values

Grades	O (90% and Above)	A (80% and Above)	B (70% and Above)	C (60% and Above)	D (50% and Above)	E (40% and Above)	F (Fail)
Ranking	0.0	0.1	0.2	0.3	0.4	0.5	0.9

Based on this general relative ranking mechanism and its associated penalty values, the relativeranking for two-class problem can be fixed as shown in Table 5.

Table 5. Relative Penalty Value for Two-Class

Grades	**pass**	**fail**
Ranking	0.0	0.9

Using this relative penalty values, the degree of misclassification cost matrix for two- class data set can derived from Table 1 as follows:

Table 6. Matrix representing degree of Misclassification Cost

			PredictedResults	
			pass	**fail**
			0.0	0.9
True Results	**pass**	0.0	**0**	**-0.9**
	fail	0.9	**0.9**	**0**

The final cost matrix calculated from the degree of misclassification using equation (4.3),with m = 0.9 and S = 100 is shown in Table 7. Since the distance from the class/grade (“***pass*** *” or “* ***fail*** *”)* itself was zero, the values of all diagonal elements became zero.

The misclassification cost for the each of the classifier based can be easily obtained by multiplying cost matrix (Table 7) with the corresponding confusion matrices of five classifiers – j48, DT, BayesNet, NaiveBayes and MLP.

Table 7. Cost Matrix for Two-Class

		Predicted Results	
		pass	**fail**
True Results	**pass**	0	30
	fail	50	0

It is noted that, prior to computing misclassification costs, each cell value of the confusion matrices is normalized by dividing the cell value by the total number of the instancesin that row. With the computed misclassification cost measure values, the performance of classifiers can be ranked.

Table 8. Performance Evaluation Results of Filter-Based Classifiers for Two-Class Student data set

Classifiers	Based on Misclassification Cost Measure		Based on Accuracy Measure	
	Cost	Ranking	Accuracy	Ranking
BayesNet-CFS	25.54592	18	85.2261	13
BayesNet-CHI	27.0665	21	85.0535	14
BayesNet-CSS	27.59583	22	85.2261	13
BayesNet-FSS	24.51467	15	85.2261	13
BayesNet-GAR	29.30358	24	85.0535	14
BayesNet-ING	29.30358	24	85.0535	14
DT-CFS	27.87417	23	85.2951	12
DT-CHI	24.51467	15	84.1042	16
DT-CSS	25.60515	19	83.8454	17
DT-FSS	24.43254	13	83.2068	20
DT-GAR	24.05142	11	83.7936	18
DT-ING	24.51467	15	83.5002	19
J48-CFS	24.06144	12	85.2261	13
J48-CHI	15.66173	9	88.0566	7
J48-CSS	15.43349	7	88.2292	6
J48-FSS	15.13625	5	90.0242	5
J48-GAR	15.33592	6	87.3317	9
J48-ING	15.65809	8	87.6942	8
NaiveBayes-CFS	26.83961	20	84.3804	15
NaiveBayes-CHI	24.69793	17	77.3731	24
NaiveBayes-CSS	25.23449	18	80.7732	21
NaiveBayes-FSS	24.55009	16	75.8026	25
NaiveBayes-GAR	24.49796	14	80.2554	22
NaiveBayes-ING	24.69793	17	77.9945	23
MLP-CFS	21.82812	10	86.3652	11
MLP-CHI	11.84857	4	97.5837	2
MLP-CSS	9.863847	2	98.5847	1
MLP-FSS	4.338674	1	87.1246	10
MLP-GAR	10.03112	3	93.3379	4
MLP-ING	10.03112	3	95.7887	3

Table 8 summarizes the misclassification costs and corresponding ranking for all of the five classifiers with feature subsets drawn from six filtered feature subset evaluation methods considered. For better understanding, we also included the overall predictive accuracy of the all the said filter-based classifiers mentioned in Table 4.4 and ranked them according to their predictive rate. On examining the performance evaluation results, we observed that the misclassification cost for the classifiers considered were different, with MLP-FSS being lowest and the BayesNet-ING being highest in the misclassification cost. Further more, for this two-class student data set, both misclassification cost and predictive accuracy measure were similar for MLP and J48 filtered classifiers.

Similar performance evaluation results were obtained for both wrapper-based and hybrid-based classifiers and they are presented in Table 9 and Table 10.

Table 9. Performance Evaluation Results of Wrapper-Based Classifiers for Two- Class Student data set

Classifiers	**Based on Misclassification cost Measure**		**Based on Accuracy Measure**	
	Cost	**Ranking**	**Accuracy**	**Ranking**
BayesNet-FFS	24.51467	5	85.2261	5
BayesNet-NB-BF	47.59255	10	85.2261	6
DT-FFS	24.43254	4	83.2068	9
DT-NB-BF	40.47911	8	84.1733	7
J48-FFS	15.13625	3	90.0242	2
J48-NB-BF	34.87157	7	87.7632	3
NaiveBayes-FFS	24.55009	6	75.8026	10
NB-NB-BF	41.34782	9	83.6382	8
MLP-FFS	4.338674	1	87.1246	4
MLP-NB-BF	7.824321	2	97.1004	1

From Table 9, we observed that, MLP-FSS had the lowest misclassification cost and BayesNet-NB-BF method had the highest misclassification cost measure. More over, the ranked values were mismatched with respect to cost as well as accuracyamong the wrapper-based classifiers.

Table 10. Performance Evaluation Results of Hybrid-Based Classifiers for Two- Class Student data set

Classifiers	**Based on Misclassificationcost Measure**		**Based on Accuracy Measure**	
	Cost	**Ranking**	**Accuracy**	**Ranking**
BayesNet-F1M-CFS-17	49.74674	30	85.0535	18
BayesNet-F1M-CHI-10	49.74674	30	85.0535	18
BayesNet-F1M-ING-13	49.74674	30	85.0535	18
BayesNet-FFS	24.51467	7	85.2261	16
BayesNet-ROC-CFS-9	48.89114	29	85.1053	17
BayesNet-ROC-CHI-9	49.74674	30	85.0535	18
BayesNet-ROC-ING-6	50	31	85.2261	16
DT-F1M-CFS-17	41.83099	27	83.6037	22
DT-F1M-CHI-10	41.56463	23	83.9144	19
DT-F1M-ING-13	41.58754	24	83.552	23
DT-FFS	24.43254	6	83.2068	26
DT-ROC-CFS-9	41.64595	26	83.5347	24

DT-ROC-CHI-9	41.24688	20	83.7763	20
DT-ROC-ING-6	46.46919	28	85.2951	14
J48-F1M-CFS-17	31.47625	10	88.4881	6
J48-F1M-CHI-10	40.56827	19	86.7449	12
J48-F1M-ING-13	37.81174	16	87.1419	9
J48-FFS	15.13625	3	90.0242	4
J48-ROC-CFS-9	41.33977	22	86.486	13
J48-ROC-CHI-9	41.28275	21	86.7967	11
J48-ROC-ING-6	50	31	85.2261	15
JNB-ROC-ING-6	41.63006	25	79.8412	30
NB-F1M-CFS-17	32.79291	11	81.9296	28
NB-F1M-CHI-10	34.77926	14	80.1174	29
NB-F1M-ING-13	33.53308	12	75.8026	31
NB-FFS	24.55009	8	83.621	21
NB-ROC-CFS-9	39.36513	18	82.3438	27
NB-ROC-CHI-9	35.94185	15	83.4311	25
MLP-F1M-CFS-17	12.30193	2	94.7877	1
MLP-F1M-CHI-10	25.62715	9	91.1288	3
MLP-F1M-ING-13	15.53182	4	94.2354	2
MLP-FFS	4.338674	1	87.1246	10
MLP-ROC-CFS-9	33.89446	13	88.1602	7
MLP-ROC-CHI-9	21.65621	5	89.9206	5
MLP-ROC-ING-6	39.27107	17	87.1591	8

By comparing the performance evaluation results of hybrid-based classifiers on both cost measure and predictive accuracy, the ranks of MLP-ROC-CFS-17, J48-FSS, MLP-F1M-ING-13 and MLP-ROC-CHI-9 were similar.

IV. CONCLUSION

The academic achievement of higher secondary school education in India is a turning point in the life of any student, as it serves as a very important link between the higher and higher secondary education of students. But, there are determinants like demographic, academic and socio-economic factors of students that restrict the students' performance. This necessitates the need for some forecasting systems to predict the academic performance of students at plus two examinations. This is an attempt made first time in this aspect, which is mainly devoted to design and develop a prediction model by taking into account variables pertaining to the Indian society, for Indian educational system. Wide literature review on academic performance of students and its prediction by using performance models was carried out. But, it was noticed that limited research investigations have been executed not only on the factors that are influencing the academic performance of the students at high school/ higher secondary level but also on the prediction of the academic performance of the students using different classification algorithm in data mining. So in this paper present the evaluation of student performance prediction models with two-class using data mining approach.

V. REFERENCES

[1] AI-Radaideh, Q. A., AI-Shawakfa, E. M. and AI-Najjar, M. I. (2006), “Mining student data using decision trees,” International Arab Conference on

Information Technology (ACIT'2006) December 19-21, 2006, Yarmouk University, Jordan.
[2] Almuallim, H. and Dietterich, T.G. (1994), “Learning Boolean concepts in the presence of many irrelevant features,” Artificial Intelligence, 69 (1-2), 279-305.
[3] Arlington (2006), “School policy and practice: Taking on childhood obesity,” American Association of School Administrators, pp.8.
[4] Baesens, B., Setiono, R., Mues, C. and Vanthienen, J. (2003), “Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation,” Management Science, 49(3), 312-329.
[5] Bekele, R. and Menzel, W .(2005), “A Bayesian Approach to Predict Performance of a Student (BAPPS): A Case with Ethiopian Students,” Proceedings of International Conference on Artificial Intelligence and Applications, 453 (191).
[6] Bell, D.A. and Wang, H. (2000), “A formalism for relevance and its application in feature subset selection,” Machine Learning, 41(2), 175–195.
[7] Ben-Bassat, M. (1982), “Pattern recognition and reduction of dimensionality,” Hand book of statistics-II, pp. 773 -791, North Holland.
[8] Brachman, R. and Anand, T. (1996), “The Process of Knowledge Discovery in Databases”, Advances in Knowledge Discovery and Data Mining, AAAI/The MIT Press.
[9] Bradley, A. P. (1997), “The use of the area under the ROC curve in the evaluation of machine learning algorithms,” Pattern Recognition, 30(7), 1145-1159.
[10] Bray, M. (2007), “The shadow education system: Private tutoring and its implications for planners,” UNESCO, PARIS, France:
[11] Brazdil, P. and Gama, J. (1994), “Characterizing the applicability of classification algorithms using meta-level learning,” Proceedings of the European conference on Machine Learning (ECML-94), Springer-Verlag.