



World Scientific News

An International Scientific Journal

WSN 113 (2018) 185-193

EISSN 2392-2192

Comparative Study of Techniques Used in Prediction of Student Performance

Minakshi Chauhan and Varsha Gupta*

KIET Group of Institution, Ghaziabad, India

*E-mail address: varsha.gupta@kiet.edu

ABSTRACT

Providing high quality education is a major concern for higher educational institutions. The quality of education in higher institutions can be assessed by the teaching and learning process. The quality of the teaching learning process depends on the performance of instructor as well as performance of students involved. Analysis and prediction of student performance is key step to identify the poor academic performance. On the basis of prediction, the corrective actions must be taken to improve performance of students and enhance the quality of education system. In this study we surveyed the techniques commonly used to predict the performance of students and also analysed the factors affecting the student academic performance.

Keywords: Educational Data Mining, Data Mining Techniques, Fuzzy Logic, Classification, Clustering

1. INTRODUCTION

Higher institutions are producing a large amount of student related data every year. Educational data mining provides a tool to analyse this data and extract valuable information from it. Then conclusion can be drawn from this mined data which will help to increase success rate of students. The process of EDM educational data mining may consists of the following steps:

1. Data collection
2. Data pre-processing
3. Data analysis
4. Prediction
5. Recommendation:

Normally the data are collected from higher institutions. Then the collected data are pre-processed if needed and the parameters are selected which are helpful in prediction based on the techniques opted. After predicting the performance appropriate support can be recommended to student so that they can improve their subsequent academic performance. Some researchers also predicted critical courses in which student face difficulties [2]. Predicting critical courses based on the previous data may help to provide support from the starting to get better performance in that particular course.

The students' performance can be predicted on the basis of courses as well as laboratory performance [3]. The researcher concluded that the understanding level of other subjects like English and Mathematics, affects the performance of student in programming courses. The author also identified the students with poor performance so that dropouts can be reduced to some extent [4]. Day by day different institutions generate huge amount of student related data which leads the problem of data storage and data analysis.

The authors identified data storage methods opted by different institutions and discussed a solution to analyse the data [5]. The data can be managed and processed in distributed and heterogeneous environment. Many researchers used different data mining techniques to extract valuable information from this large amount of student related data. In this paper, we studied and presented a comparative analysis of those techniques.

2. TECHNIQUES USED

Researcher used different approaches to predict academic performance of students. These approaches were based on fuzzy logic, data mining techniques [6], and machine learning techniques etc. In this study, we have discussed proposed predicting techniques by different researchers into following three categories: Fuzzy logic, Data mining, Hybrid.

2. 1. Fuzzy logic

Fuzzy logic is used to address and analyze the problems which include a degree of uncertainty. The general fuzzy technique used in student performance prediction can be broadly described in following three steps:

1. Fuzzify input parameters and output parameters.
2. Determine rules base and inference method.
3. Defuzzify output parameters.

The comparative analysis of fuzzy based techniques is shown in Table 1. Most of researchers considered students academic performance specially marks or score to design the input for proposed approaches.

Table 1. Comparison of fuzzy logic based approaches

S. No.	Paper	Parameters used	Membership function used	Data Set size	Rule Set
1	[3]	Exam1 Exam2	Triangle	20	25
2	[5]	Attendance, Internal Exam Evaluation and Term Work Evaluation	Triangle and Trapezoidal	NA	27
3	[7]	Semester1 Semester2	Triangle	20	25
4	[8]	Atten3, Atten4, Atten5, Marks3, and Marks4	Gaussian, Trapezoidal, and Sigmoidal	120	30
5	[9]	Exampaper1 Exampaper2	Trapezoidal	20	36

It has been clearly observed that mostly researchers used small data set size and rule base. The method based on fuzzy concept was better than the classical method to capture the progress of student performance in subsequent semester [7].

Table 2. Comparison of data mining techniques

S. No.	Approach used	Parameters used	Data Set size	Accuracy (%)
1	ID3 Decision Tree Induction Algorithm [2]	Student ID, Graduation GPA, HSC Score, Aptitude Test, Educational Attainment Test, Courses	100	80%
2	Classification Association Rules Mining (CARM) [4]	Graduation year, GPA, student ID, course code	203 records	NA

3	K-mean Cluster Algorithm [11]	gender, national origin, Parental job, GPA,	300 students grouped into 3 clusters	Variable
4	Decision Tree [12]	Branch HSC Percent HSC Details SSC Percent SSC Details Category Gender Living Location SSC Board Parents Occupation	346	90.7
5	K-Means Clustering Algorithm with the Euclidean Distance Measure [13]	Student score	79 students in 9 courses	Variable
6	Decision Tree-ID3Algorithm [14]	Class Test Grade, Seminar Performance Assignment, General Proficiency, Attendance, Lab Work, Semester Marks	50	NA

2. 2. Data mining techniques

Data mining techniques are extensively being used to analyse data and discover knowledge in different fields such as medical, engineering, marketing, web and education etc. In this section we included data mining techniques such as Decision tree, Neural Network, Naïve Bayes, K-nearest neighbour, Support vector machine etc. used to predict student academic performance. The techniques can be used in classification, association rules and clustering. Classification method in data mining categorises the items or objects into predefined classes and predicts the category of new items or objects based on those classes. For example, Classification can be used to classify the student in two classes such as “at risk” and “safe” in terms of their performance [10]. Clustering method in data mining tries to form groups or clusters of objects based on their relationship. For example, Clustering can be used to categorize students into three clusters low performance, average performance and smart performance [11].

Comparative study of data mining techniques used in prediction has been given in Table 2 based on the attributes such as approach used, parameters used, performance measure used, data size taken, accuracy claimed. During the survey it has been observed that most of researchers used decision tree and k-mean clustering for predicting student performance.

2. 3. Hybrid models

In this category we considered the approaches combining two or more algorithms. The comparative analysis of approaches based on the attributes such as approach used, parameters used, performance measure used, data size taken, accuracy claimed has been shown in Table 3. Most of researchers combined decision tree and artificial neural network.

Table 3. Comparison of approaches combing two or more data mining techniques.

S. No.	Approach used	Parameters used	Data Set size	Accuracy (%)
1	Decision Tree and FGA[10]	Internal Sessional AdmScore	120 UG students and 48 PG students	NA
2	Three ANFIS models [15]	5 Compulsory Subject Marks	100	NA
3	ANN, Decision Tree and Linear Regression [16]	CGPA	NA	More than 80 %
4	Decision Tree, Smooth SVM, Kernel K-Means Clustering [17]	Interest, Study Behaviour, Engage Time, Race, Believe, Family Support, CGPA, Gender, Religion	100 students	97.3%

3. COMMON FACTORS AFFECTING STUDENT PERFORMANCE

To apply any model on student historical data, first it is required to decide the attributes on which the prediction will be done [18]. Attribute selection methods can be applied to find important attributes applicable to particular technique opted. These attributes can be personal, family background related, academic/education related, and behaviour or geographical related.

We categorized parameters selected by different researchers into five categories namely Academic, Behavioural, Personal, Geographical, and Family.

The graphical representation of number of parameters opted versus approach proposed is shown in Figure 1. This representation clearly shows that most of researchers focussed on academic attributes in prediction. Along with the academic parameters, student behavioural parameters affect their final academic performance [17]. It has been observed that the family background and parents occupation has huge impact on academic performance of their child.

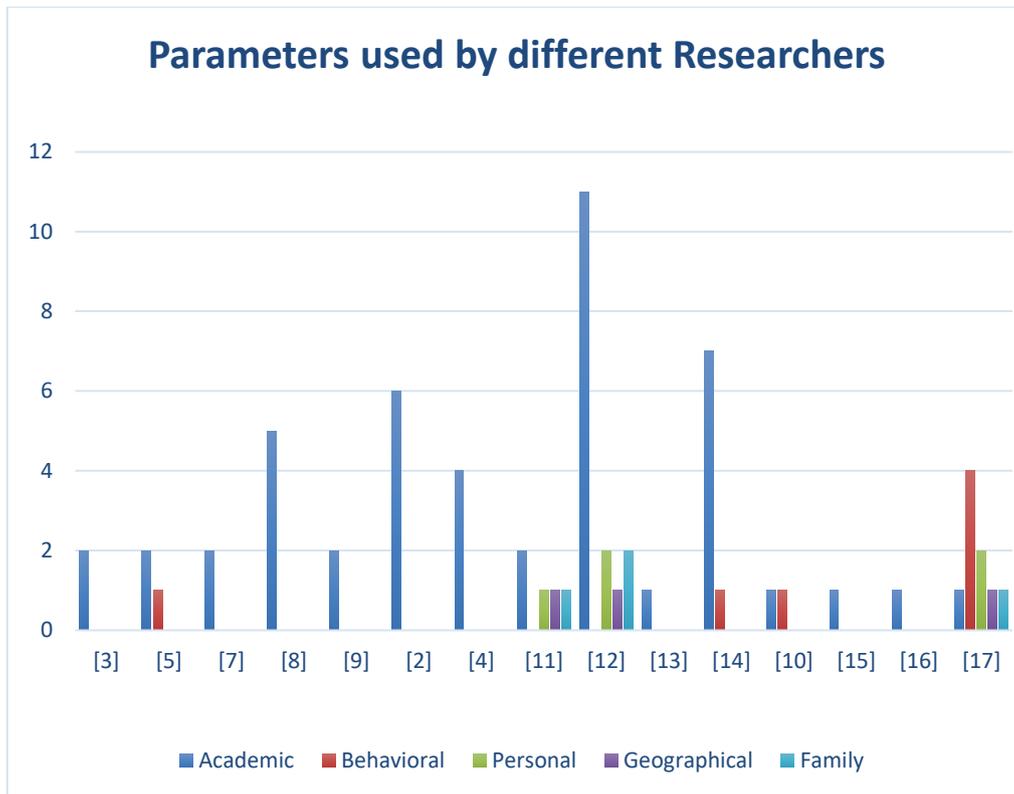


Figure 1. Number of parameters used by different approaches

4. PERFORMANCE MEASURES

Table 4. Performance Measure methods

S. No.	Paper	Performance Measure
1	[3]	compared with classical evaluating method
2	[5]	NA
3	[7]	compared with classical evaluating method and Fuzzy1

4	[8]	RMSE
5	[9]	compared with classical evaluating method and Fuzzy1
6	[2]	accuracy, precision, and recall
7	[4]	support, confidence, and accuracy rate
8	[11]	NA
9	[12]	10-fold cross-validation
10	[13]	NA
11	[14]	NA
12	[10]	NA
13	[15]	Root Mean Square Error (RMSE)
14	[16]	square root of average squared error (RASE)
15	[17]	Accuracy

Performance measures are used to evaluate the prediction correctness of any proposed approach. The common performance measures mostly used are statistical methods which include RMSE, RASE, precision, and recall, support, confidence, and accuracy rate. Precision, recall and accuracy rate are common performance measures used to evaluate the data mining approaches. Some of researchers use bench mark models or previous approaches to prove that the proposed approach is better. Performance Measure evaluation for different approaches has been shown in Table 4. For some proposed approaches, Performance Measure method was not given. So NA is mentioned.

5. CONCLUSION

This study has surveyed the techniques applied to predict student performance and categorized majorly into three categories namely fuzzy logic, data mining techniques and hybrid. The techniques under these three categories have been compared based on the methodology used, size of data set operated on, and parameters taken and so on. Further this study also identified the common factors affecting student performance most. During the study it has been observed that historical data related to the marks plays an important role in predicting the future academic performance of students. Some researchers also focused on the attributes related to the family background of the student which have great influence on the performance.

References

- [1] Goga, M., Kuyoro, S. and Goga, N., A recommender for improving the student academic performance. *Procedia - Social and Behavioral Sciences*, 180, pp. 1481-1488, (2015).
- [2] Altujjar, Y., Altamimi, W., Al-Turaiki, I. and Al-Razgan, M., Predicting Critical Courses Affecting Students Performance: A Case Study. *Procedia Computer Science*, 82, pp. 65-71, (2016).
- [3] Gokmen, G., Akinci, T.Ç., Tektaş, M., Onat, N., Kocyigit, G. and Tektaş, N., Evaluation of student performance in laboratory applications using fuzzy logic. *Procedia - Social and Behavioral Sciences*, 2(2), pp. 902-909, (2010).
- [4] Badr, G., Algobail, A., Almutairi, H. and Almutery, M., Predicting students' performance in university courses: a case study and tool in KSU mathematics department. *Procedia Computer Science*, 82, pp. 80-89, (2016).
- [5] Patel, S., Sajja, P. and Patel, A., Fuzzy logic based expert system for students performance evaluation in data grid environment. *International Journal of Scientific & Engineering Research* 5(1), (2014).
- [6] Shahiri, A.M. and Husain, W., A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, pp. 414-422, (2015).
- [7] Yadav, R.S. and Singh, V.P., Modeling academic performance evaluation using soft computing techniques: A fuzzy logic approach. *International Journal on Computer Science and Engineering*, 3(2), pp. 676-686, (2011).
- [8] Chauhan, M., Rastogi, A. and Kapoor, U., Fuzzy Based Prediction of Student's Performance in University Exam: A Case Study, *International Journal of Research in Engineering, IT and Social Sciences*, Volume 08, Special Issue, Page 138-145, (2018).
- [9] Sakthivel, E., Kannan, K.S. and Arumugam, S., Optimized evaluation of students performances using fuzzy logic. *International Journal of Scientific & Engineering Research*, 4(9), pp. 1128-1133, (2013).
- [10] Hamsa, H., Indiradevi, S. and Kizhakkethottam, J. J., Student Academic Performance prediction model using decision tree and fuzzy genetic algorithm, *Procedia Technology*, Vol. 25, 326-332, (2016).
- [11] Alfiani, A.P. and Wulandari, F.A., Mapping student's performance based on data mining approach (a case study). *Agriculture and Agricultural Science Procedia*, 3, pp. 173-177, (2015).
- [12] Kabra, R.R. and Bichkar, R.S., Performance prediction of engineering students using decision trees. *International Journal of Computer Applications*, 36(11), pp. 8-12, (2011).
- [13] Oyelade, O.J., Oladipupo, O.O. and Obagbuwa, I.C., Application of k Means Clustering algorithm for prediction of Students Academic Performance. *International Journal of Computer Science and Information Security*, Vol. 7, No. 1, 292-295, 2010 arXiv preprint arXiv:1002.2425 (2010).

- [14] Baradwaj, B.K. and Pal, S., Mining educational data to analyze students' performance. *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 6, 63-69, 2011 arXiv preprint arXiv:1201.3417
- [15] Altaher, A. and BaRukab, O., Prediction of Student's Academic Performance Based on Adaptive Neuro-Fuzzy Inference. *International Journal of Computer Science and Network Security* 17(1), p. 165, (2017).
- [16] Ibrahim, Z. and Rusli, D., Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression. In 21st Annual SAS Malaysia Forum, 5th September, (2007).
- [17] Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S. and Wani, E., Prediction of student academic performance by an application of data mining techniques. In International Conference on Management and Artificial Intelligence IPEDR Vol. 6, No. 1, Pp. 110-114, (2011).
- [18] Borkar, S. and Rajeswari, K., Predicting students academic performance using education data mining. *International Journal of Computer Science and Mobile Computing*, 2(7), pp. 273-279, (2013).