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.
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

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Paper</th>
<th>Parameters used</th>
<th>Membership function used</th>
<th>Data Set size</th>
<th>Rule Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[3]</td>
<td>Exam1, Exam2</td>
<td>Triangle</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>[7]</td>
<td>Semester1, Semester2</td>
<td>Triangle</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>[8]</td>
<td>Atten3, Atten4, Atten5, Marks3, and Marks4</td>
<td>Gaussian, Trapezoidal, and Sigmoidal</td>
<td>120</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>[9]</td>
<td>Exampaper1, Exampaper2</td>
<td>Trapezoidal</td>
<td>20</td>
<td>36</td>
</tr>
</tbody>
</table>

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

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

-187-
### 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].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Decision Tree [12]</td>
<td>Branch HSC Percent HSC Details SSC Percent SSC Details Category Gender Living Location SSC Board Parents Occupation</td>
<td>346</td>
</tr>
<tr>
<td>5</td>
<td>K-Means Clustering Algorithm with the Euclidean Distance Measure [13]</td>
<td>Student score 79 students in 9 courses</td>
<td>Variable</td>
</tr>
<tr>
<td>6</td>
<td>Decision Tree-ID3Algorithm [14]</td>
<td>Class Test Grade, Seminar Performance Assignment, General Proficiency, Attendance, Lab Work, Semester Marks</td>
<td>50</td>
</tr>
</tbody>
</table>

-188-
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>
<thead>
<tr>
<th>S. No.</th>
<th>Approach used</th>
<th>Parameters used</th>
<th>Data Set size</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Decision Tree and FGA[10]</td>
<td>Internal Sessional AdmScore</td>
<td>120 UG students and 48 PG students</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>Three ANFIS models [15]</td>
<td>5 Compulsory Subject Marks</td>
<td>100</td>
<td>NA</td>
</tr>
<tr>
<td>3</td>
<td>ANN, Decision Tree and Linear Regression [16]</td>
<td>CGPA</td>
<td>NA</td>
<td>More than 80 %</td>
</tr>
<tr>
<td>4</td>
<td>Decision Tree, Smooth SVM, Kernel K-Means Clustering [17]</td>
<td>Interest, Study Behaviour, Engage Time, Race, Believe, Family Support, CGPA, Gender, Religion</td>
<td>100 students</td>
<td>97.3%</td>
</tr>
</tbody>
</table>

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.

![Parameters used by different Researchers](image)

**Figure 1.** Number of parameters used by different approaches

4. PERFORMANCE MEASURES

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Paper</th>
<th>Performance Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[3]</td>
<td>compared with classical evaluating method</td>
</tr>
<tr>
<td>2</td>
<td>[5]</td>
<td>NA</td>
</tr>
<tr>
<td>3</td>
<td>[7]</td>
<td>compared with classical evaluating method and Fuzzy1</td>
</tr>
</tbody>
</table>
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


