Analysis of Performance Parameter using Data Mining Techniques

Archana Bhardwaj a, Mamta Bhusry (Guide?) b *

a Student: Ajay Kumar Garg Engineering College Ghaziabad, Uttar Pradesh, India, archanaakgc@gmail.com
b Asst. Prof. (CSE): Ajay Kumar Garg Engineering College Ghaziabad, Uttar Pradesh, India, bhusrymb@yahoo.co.in

Abstract

Educational data mining is the emerging research field of Data Mining. The education system generates vast data on student, faculty, accounts and others in their database with the help of modern computing technology. This paper applies data mining techniques to help the management of the education system to make better decision-making strategies to improve the performance of students. Not only the students are responsible for the development of the education system, as the data mining techniques are applied to the student domain in the same way data mining techniques can be applied to the faculty domain to enhance their performance.

Keywords: Data Mining, Performance Analysis, Education System, Decision Making

1. Introduction

Data Mining is the process of extracting useful knowledge from the huge amount of data by combining some techniques like statistical and artificial intelligence [10]. Data mining uses a mathematical algorithm for calculating the probability of the future event [6]. The goal of data mining is to discover a frequent pattern or structure and interesting correlation from large data repositories [3]. Data mining that pulled out information from huge data set have become popular in different field such as marketing, fraud detection and medical etc.[8],[1]. One of the emerging research areas in education framework is Educational Data Mining. Educational Data Mining can be defined as the application of data mining techniques which uses the education data[11],[12]). The main aim of education data mining is to develop better decision making strategies to enhance the performance with respect to student, faculty and the organization as a whole.

This paper analyzes the performance of the education system to better understand the student performance using the different data mining techniques in the student domain. Performance analysis provides feedback to performers trying to get a chance in performance. Researches show that on average the performer can only recall 30% of the performance. So, for the remaining 70% the performance analysis is necessary[16]. For performing analysis data mining tool is required. The process of performance analysis involves three basic steps i.e. data collection, data selection and transformation and data visualization[9].

Till now most of the research is going in the field of Educational Data Mining on the student domain to know about their performance and interest in different courses etc. using different data mining techniques such as association, classification and clustering and so on using data mining tools. For future work these Data Mining techniques can be applied to the faculty domain by taking the various parameters of the education system to improve and enhance their performance.

The whole paper is organized as: In the next section the overview of related work in which educational data mining on student domain is discussed, then that work is summarized in tabular form. In section 3 proposed work brings up where educational data mining is presented with the help of proposed model. Finally the paper is concluded in section 4 with some challenges and work in this area.

2. Literature Survey

Educational Data mining is an emerging research area in the field of data mining. One can use the data mining
techniques to enhance the understanding of learning process to focus on discovering, extracting, and evaluating variable related to the learning process of studying (Remero and Venutra, 2005, Abdullah et.al, 2011). Researchers have contributed in various ways to perform analysis using data mining techniques.

Delavari et.al (2005) [7] presented a Model called DM_EDU which is used as one of the data mining application. This Model is applied to the student the dataset to predict which student is more likely to perform well in that specific course or those who are less likely to be successful in it, so that faculty, can provide extra basic courses, skill classes, academic counseling. The model DM_EDU is applied to the dataset of 841 students. Apriori algorithm is used for data selection and transformation for this model.

The final result of such application is improving the quality of management system by taking better decision through data mining.

Shyamala et.al (2006) [13] presented a Model to improve the student performance on the basis of different parameter such as assignment marks, attendance, test marks etc. by applying data mining techniques.

The goal of the paper is to make the predictions about the student performance as well as to find similar pattern from the collected data. The test data are taken from Dr.Ambedkar Govt.College, Chennai from zoology course of 180 students. The software used for this work is Clementine and an algorithm applied is C5.0 for this model.

The paper concluded that which student is less likely to perform well in the specific course after getting the result the college mat take necessary decisions.

Buldu and Ucgn (2010) [5] presents a sample study on Data Mining application on student data, in which he has used an Apriori Algorithm in educational field and identified the results. The author also made some rules about student course. To perform the work author design “School Automation Software” to reveal association rules from database. Apriori algorithm is used in a student data set of Istanbul Eyup L.M.K.B Vocational Commerce High School and result are observed and analyzed.

The paper concludes that the student who is unsuccessful in numerous courses in their previous class will be unsuccessful in their present class too. So according to the result the student can be helped in choosing their profession and their concern field.

Ayesha et.al (2010) [2] uses the data mining process upon the student dataset through K-means clustering algorithm to predict the student learning activities. This may improve student performance by reducing the failing ratio and take appropriate steps to improve the quality of education. Data is collected from computer science department of the University of Agriculture of 120 students and perform their work using visual studio 2005 and SQL server 2005.

The paper concludes that by applying data mining techniques on student database one can take appropriate step to improve the quality of education.

Zengin et.al (2011) [15] presents a sample study in this paper the author analyze the data using descriptive statistics and data mining techniques. Author uses the Delphi 2009 programming language on Microsoft SQL server 2008 on the dataset “computer Self-efficiency scale” of educational science.

The paper carried out that it is not essential to use a data mining techniques for analyzing data, instead of data mining package one can use multiple methods to obtain the best results.

Bhardwaj and Pal (2011) [4] use the data mining methodology to predict the performance of the student. Here the author uses the classification technique to evaluate the performance and identify the students who need the social attention to reduce the fail ratio. Data is gathered from VBS Purvanchal University Jaunpur of 50 students. After performing the data processing techniques decision tree technique is applied by using ID3 algorithm.

The paper concludes the performance of student in their end semester exams and helps to manage the drop out student and also manage the students who need special attention, so that teacher may provide appropriate counseling to them.

Weng (2011) [14] proposed a new algorithm name as FARM to discover fuzzy rare item sets and applied on the educational datasets. This algorithm is applied on the datasets of college of management of 62875 students and the result shows the feasibility of the proposed FARM algorithm.

From the experiment result, it concludes that the proposed algorithm is able to discover interesting and valuable patterns from education data.

The limitation of all the above studied paper is that they all have taken the minimum amount of data and have not mentioned whether they are scalable or not.

Table 1 shows the literature survey in summarized form. In which all the main points are taken to describe the related work.

3. Proposed Work

Most of the work in the education system has been carried out for analyzing student performance using data mining techniques. Measurement and analysis of faculty performance is equally important from the point of view of faculty development and reorganization. However not much has been done in this area. We propose to apply the data mining technique in faculty performance evaluation in the education system. The proposed work would be done from the observation of faculty job profile and discussion with senior faculties in the education system. In our work we are going to apply data mining techniques on the faculty of computer science and engineering department of engineering college of ABC the name is not disclosed for purpose of anonymity.

The objective is to perform an analysis considering a
number of parameters for the derivation of performance prediction needed for faculty assessment, monitoring and evaluation. The aim is to estimate the quality, productivity and potential of the faculty across various parameters. The analysis of the faculty depends on many factors such as student’s feedback, teaching method, seminars, paper and book published assessment and research activity.

A holistic Faculty Evaluation using data mining technique is proposed on the giving parameters:
- Results in courses taught by faculty
- Research activity
- Student Feedback
- Seminar, workshop and conferences organized by

The proposed work is to perform an analysis of faculty performance considering above all parameters which affect the impact of management decision, the standard of teaching, satisfaction, growth and decline in the performance. The statistical analysis and data mining techniques which help in the extraction of useful pattern from the college database will be used for extracting useful patterns in faculty performance.

<table>
<thead>
<tr>
<th>GOAL</th>
<th>DATASET</th>
<th>DATASIZE</th>
<th>DATA PROCESSING</th>
<th>ALGORITHM</th>
<th>SOFTWARE</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper</td>
<td>Has been a motivation Towards Enhancing the proposed modal DM_EDU that is used as the application of data mining in higher education system</td>
<td>Data set Multimedia in Malaysia</td>
<td>Data selection And transformation</td>
<td>Apriori algorithm, CRISP methodology</td>
<td>Delavari et. al, 2005</td>
<td></td>
</tr>
<tr>
<td>To predict which student is more likely to perform well in the specific course or those who are less likely to be successful in it</td>
<td>Data set from Dr. Ambedkar govt. College Chennai</td>
<td>180 students</td>
<td>C5.0</td>
<td>Clementine</td>
<td>Shyamala Rajagopalan, 2006</td>
<td></td>
</tr>
<tr>
<td>To realize data mining with Apriori algorithm of student data</td>
<td>Eyup L.M.K.B Vocational Commerce high school</td>
<td>28 students</td>
<td>Selection and transformation</td>
<td>Apriori ago</td>
<td>Buldu &amp; Ucgun, 2010</td>
<td></td>
</tr>
<tr>
<td>Performance during academic career using K-means and decision tree in the educational system</td>
<td>Computer Science department of university Of Agriculture</td>
<td>120 students</td>
<td>Data selection And transformation</td>
<td>K-means clustering</td>
<td>Visual Studio 2005, SQL server 2005</td>
<td></td>
</tr>
<tr>
<td>To present a sample study analyzing data gathered for educational study using data mining techniques</td>
<td>Repository of student of Goziosmanp Ose University</td>
<td>531 students</td>
<td>Data transformation</td>
<td>Decision tree, Dependency w, clustering</td>
<td>Delphi2009 Programming language on Microsoft SQL Server 2005</td>
<td></td>
</tr>
<tr>
<td>Investigates the accuracy of the decision tree technique for predicting student Performance</td>
<td>VBS paravanchal University Jaunpur</td>
<td>90 students</td>
<td>Selection and Transformation</td>
<td>Decision tree,&lt;ID3&gt;</td>
<td>Rhandwaj &amp; Pal, 2011</td>
<td></td>
</tr>
<tr>
<td>Propose an Apriori based Mining Approach Fuzzy Apriori rare itemset mining (FARIM)</td>
<td>Data set Obtained From Computer Center University</td>
<td>52875</td>
<td>Fuzzy Apriori Algo, Fuzzy Apriori rare, FARIM</td>
<td>San Java J2SDK1.3.1</td>
<td>Weng, 2011</td>
<td></td>
</tr>
</tbody>
</table>
The ongoing research on the faculty performance has allowed researchers to increase the data size and consider all the possible factors engaged in the faculty performance within the organization. We are combining more constraint which can affect the faculty performance.

Fig. 1 shows the procedure to achieve the valuable parameter of faculty in Educational Data Mining. This is divided in five parts: Data collection, Data selection, Data preparation, Tools and Techniques, Result Analysis and last one is Decision Making. Enable

3.1 Data Collection

Data collection is the process of gathering and measuring information of interest that enables to answer stated research questions and evaluate outcomes. For our work the data regarding following parameters will be collected from the Engineering College ABC:
- Results in courses taught by faculty
- Research activity
- Student Feedback
- Seminar, workshop and conferences organized by the faculty
- Initiatives for self improvement and career development.
- Administrative responsibility carried out by the faculty.

Data about student feedback and Results taught by faculty will be collected from Engineering College ABC. Data related to Faculty personal information such as total experience, research activity etc. will be from faculty itself.

<table>
<thead>
<tr>
<th>Field of Result marks</th>
<th>Range of marks of result data</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;90 marks</td>
<td>75-90 marks</td>
</tr>
<tr>
<td>69-74 marks</td>
<td>49-59 marks</td>
</tr>
<tr>
<td>30-44 marks</td>
<td>&lt;30 marks</td>
</tr>
<tr>
<td>absent</td>
<td>Not declared(N/D)</td>
</tr>
</tbody>
</table>

Table 2 Field of Result marks

<table>
<thead>
<tr>
<th>Different parameters of student feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage of subject material specified in the syllabus</td>
</tr>
<tr>
<td>Preparation and organization of lectures</td>
</tr>
<tr>
<td>Quality of tutorials</td>
</tr>
<tr>
<td>Uniformity in pace of Teaching during the semester</td>
</tr>
<tr>
<td>Presentation and Communication</td>
</tr>
</tbody>
</table>

3.2 Data preparation

For our work we choose Microsoft Excel to design datasets. All the information collected from data collection phase will then transform into Excel spreadsheets. The results and student feedback parameter will be stored on the basis following field as shown in Table 2 and Table 3.

3.3 Tools and Techniques

Today, in computer era different data mining tool are available in the market and each tool has its own merits and demerits. For the analysis of data of Engineering College ABC, we have concentrated on faculty data and will study the predictive analysis to look into which parameter is more effective for faculty performance using data mining clustering, classification and association techniques. For this we are going to use the IBM SPSS Statistic version 20.0 for mining and analysis activity.

IBM SPSS Statistic 20.0 is software for statistical analysis. The advantage of IBM SPSS Statistic 20.0 is that it supports any file and generates reports, charts, graphs, plots of distribution etc. In this user can enter data by using Excel spreadsheets, notepad, and Microsoft access or can create data in programming itself. SPSS can also read data files from variety of file formats including Excel, SAS, and Stata.

3.4 Result Analysis

In our further work after applying these data mining techniques the result will be analyzed according to the given input parameters.

The first objective is to analyse the semester wise performance of the faculty on the basis of result and student feedback using clustering data mining technique and
classify them in specific class as shown in Table 4.

<table>
<thead>
<tr>
<th>Class division for performance</th>
<th>Excellent</th>
<th>Very good</th>
<th>Good</th>
<th>Satisfactory</th>
</tr>
</thead>
</table>

Second objective is to analyze the performance of those faculty who are teaching the same subject in different group and accordingly put them in the priority table as shown in Table 5.

<table>
<thead>
<tr>
<th>Subject/priority</th>
<th>Priority 1</th>
<th>Priority 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the analyzed results we will define good decision making strategies for better performance of the faculty and can define which parameter is more valuable for the performance enhancement of the both faculties as well as organization.

4. Conclusion

In this paper we studied the research area in Educational Data Mining applied in the student domain to know their performance and interest and then apply appropriate decision–making strategies to improve their performance. After studying last decade papers we observe that there is a scope for improvement in Educational Data Mining. We discussed and proposed the Educational Data Mining with respect to a faculty domain to analyze the performance of faculty by taking different parameters. Our research goal is to decide which parameter is more valuable for analyzing the faculty performance. The further work is to implement the proposed work and analyze the performance parameter.

References

[16] http://www.eis2win.co.uk/Pages/Performance Analysis. Asp