A Survey on Forecasting Students Performance using EDM

C. R. Durga devi
Assistant Professor
Department of Information Technology
NGM College, Pollachi

Abstract

Education is the process by which society deliberately transmits its gathered knowledge, values, and skills from one generation to another through institutions. Educational data mining concerns with developing techniques for analyzing data and acquire knowledge from the data that comes from educational domain. Education Data Mining (EDM) is a promising discipline which has an imperative impact on predicting student’s academic performance. Student’s academic performance in educational environment is based upon numerous factors. The main goal of educational data mining is to advance the student performance in academics. This paper studies various data mining techniques like prediction, classification, clustering, association analysis that has been applied and examined to evaluate the student’s performance.

Keywords: Association Analysis, Classification, Clustering, Education Data Mining, Prediction

I. INTRODUCTION

Data Mining, also popularly known as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. Data mining is not specific to one type of media or data. Data mining should be applicable to any kind of information repository. However, algorithms and approaches may differ when applied to different types of data. Data Mining is a multidisciplinary field, encompassing areas like information technology, machine learning, statistics, pattern recognition, data retrieval, neural networks, information based systems, artificial intelligence and data visualization [12]. Applying data mining technique in educational setting is called as educational data mining. Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. EDM is a field that exploits statistical, machine-learning, and data-mining algorithms over the different types of educational data. Its main objective is to analyze these types of data in order to resolve educational research issue [1]. There are two types of data mining tasks: descriptive data mining tasks that describe the general properties of the existing data, and predictive data mining tasks that attempt to do predictions based on inference on available data. There are many popular methods in Educational Data Mining. Some of them are widely used as prediction, classification, clustering, Association Analysis.

A. Prediction

Prediction has attracted considerable attention given the potential implications of successful forecasting in a business context. There are two major types of predictions: one can either try to predict some unavailable data values or pending trends, or predict a class label for some data. The objective is forreting student’s performance

B. Classification

Classification analysis is the organization of data in given classes. It is also known as supervised classification, the classification uses given class labels to order the objects in the data collection. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects. Classification methods like decision trees, Bayesian network etc can be applied on the educational data for predicting the student’s performance in examination. This prediction will help to identify the weak students and help them to score better marks [2]. Brijesh Kumar Bhardwaj and Saurabh Pal states that Bayesian classification method is used on student database to predict the students division on the basis of previous year database. This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce failing ration and taking appropriate action at right time. [3]

C. Clustering

Similar to classification, clustering is the organization of data in classes. However, unlike classification, in clustering, class labels are unknown and it is up to the clustering algorithm to discover acceptable classes. Clustering is also called unsupervised classification, because the classification is not dictated by given class labels. M. Durairaj and C. Vijitha[10] states that Using K-
Means, one of the best clustering algorithm, we predicted the pass percentage and fail percentage of the Overall students appeared for a particular examination

D. Association Analysis:

Association analysis is the discovery of what are commonly called association rules. It studies the frequency of items occurring together in transactional databases, and based on a threshold called support, identifies the frequent item sets. Another threshold, confidence, which is the conditional probability that an item appears in a transaction when another item appears, is used to pinpoint association rules. Association rules can be used to identify Grade of the students. Rakesh Kumar Arora[1], Dr. Dharmendra Badal[11] used Apriori Algorithm to determine set of weak students in current semester by comparing the performance of students of previous semesters on the basis of marks obtained at graduate and post graduate level. The organization of the paper is presented in the next section

II. RELATED WORK

Brijesh Kumar Bhardwaj and Saurabh Pal [3] conducted study on student’s performance. Bayesian classification method was used in their study. A student database having total of 300 (226 males, 74 females) students of BCA course from five colleges who appeared in 2010 examination were the samples for this study. From their study, it is found that the students’ performance is highly dependent on their grade obtained in Senior Secondary Examination, Living Location, medium of teaching. Their study shows that academic performances of the students are not always depending on their own effort. Our investigation shows that other factors have got significant influence over students’ performance. This proposal will improve the insights over existing methods.

K.Shammuga Priya and A.V.Senthil kumar [4] carry out a study using ID3 Decision tree algorithm. In this analysis, the data set is obtained from the department of M.Sc IT 2009 to 2012 batch Hindustan College of Arts and Science, Bharathiar University Affiliated, Coimbatore .They have used several attributes like Overall Semester Marks, class test,seminar,communication skill, paper presentation,Attendance,End semester Marks. They calculate the values for Entropy, Gain, Split Information and Gain Ratio for each attribute. The attribute Overall Semester Marks has the maximum gain value, so it is the root node of the decision tree. The concept of Decision Tree which integrates with the classification technique helps to achieve the goal by extracting the discovery of knowledge from the end semester mark. The inclusion of extracurricular activities makes to gain more knowledge along with the End semester Mark. This technique is one of the ways to improve performance of the students in education.

Surjeet Kumar Yadav and Saurabh Pal [2] had done a research by collecting data from VBS Purvanchal University, Jaunpur (Uttar Pradesh) on the sampling method for Institute of Engineering and Technology for session 2010. Initially size of the data is 90. The C4.5, ID3 and CART decision tree algorithms were applied. The C4.5 decision tree algorithm can learn effective and better than other algorithms. It can learn effective predictive models from the student data accumulated from the previous years. This classifier successfully identifying the students who are likely to fail.

Umesh Kumar Pandey S. Pal [5] applied Bayesian classification method in his research. The data set used in this study was obtained from different colleges on the sampling method of computer science department of course PGDCA of session 2009-10. Initially size of the data is 600. The classifier is applied on student database to predict the students division on the basis of previous year database. This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce failing ration and taking appropriate action at right time.

Dorina Kabakchieva [6] made a comparison on performance of various classifiers. The dataset used for the research purposes includes data about students admitted to the university in three consecutive years. Several well known data mining classification algorithms, including a rule learner, a decision tree classifier, a neural network and a Nearest Neighbour classifier, are applied on the dataset. The highest accuracy is achieved for the Neural Network model (73.59%). Shafiq Aslam and Imran Ashraf [7] describe Application of some important Algorithms C4.5, SVM, EM, PageRanker, Naive Bayes, Apriori and CART in Education Data Mining.

Mahendra Tiwari, Randhir Singh, Neeraj Vimal [8] Conducted a study by using different Data mining Techniques like Association, Clustering, Classification in EDM field. The resulting Association rule depicts a sample of discovered rules from data for student with poor grade along with their support and confidence. A Rule-based classification extracts a set of rules that show relationships between attributes of the data set and the class label. It used a set of IF-THEN rules for classification. The K-means algorithm, the best one of the clustering algorithms is used in their work. They applied data mining techniques to discover knowledge, association rules. Classification rules to predict the students’ performance, as well as they clustered the students in to groups using K-Means Algorithm.Mohammed M. Abu Tair, Alaa M. El-Halees[9] performs outlier detection in students database.Two outlier methods are used which are Distance-based Approach and Density-Based Approach.

III. ANALYSIS OF DEVELOPMENTS IN EDM

We can use the data mining in educational system as: predicting drop-out student, relationship between the student university entrance examination results and their success, predicting student's academic performance, discovery of strongly related subjects.
in the undergraduate syllabi, knowledge discovery on academic achievement, classification of students’ performance in computer programming course according to learning style, investing the similarity and difference between schools.

EDM seeks for applying data mining techniques to the following educational problems
1) Providing Feedback for Supporting Instructors
2) student modeling,
3) Detecting Student Behavior
4) Predicting Student’s Performance
5) Recommendations for Students
6) grouping students
7) Constructing Courseware
8) Planning and Scheduling
9) students Social Network Analysis

The following table shows the papers that were most cited in Google scholar (citation1) and SciVerse scopus (citation2)

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Citation1</th>
<th>Citation2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data mining in education</td>
<td>74</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>Web usage mining for predicting final marks of students that use Moodle courses</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>Predicting students' performance in distance learning using machine learning techniques</td>
<td>69</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>Educational data mining: A survey and a data mining-based analysis of recent works</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>Educational data mining: A survey from 1995 to 2005</td>
<td>768</td>
<td>578</td>
</tr>
<tr>
<td>6</td>
<td>Educational Data Mining: A Review of the State of the Art</td>
<td>453</td>
<td>313</td>
</tr>
<tr>
<td>7</td>
<td>The State of Educational Data Mining in 2009: A Review and Future Visions</td>
<td>443</td>
<td>290</td>
</tr>
</tbody>
</table>

The below graph shows approximate number of full papers published in International EDM Conferences conducted in last five years. These details shows that there is a growth in EDM Research area.

Fig. 1: No. of full papers published in International EDM conferences in last five years

IV. CONCLUSION

This paper surveys the various data mining techniques like classification, clustering, Association rules that were applied in Education data mining. It also focuses on current trends and developments in Education data mining. Prediction on student academic performance is one of the most modern topics in educational data mining. The student’s performance depends on factors such as student family income, student family size, mother’s qualification, student learning behavior. The main objective of Educational Institution is to provide the best quality education to its students and to improve their behavior. Prediction of student academic performance will help the teachers to predict about student success and failure in examination. The prediction model acts like a warning system to identify potential weak students, so that the teachers can take an appropriate action towards them. Otherwise, they can give proper advices to prevent failure in the examinations. EDM plays a vital role in this area.

REFERENCES


