Applications of Data Mining Techniques in Student Performance Analysis A Survey

Nilar Aye
University of Computer Studies (Hpa-an)

Abstract - Educational Data Mining (EDM) is one of the useful applications of data mining. In EDM, there are two key research areas, one is related with students (both high school students and university students) and the other is with faculty (both teachers and institutional responsibility). A detailed research work has already been carried out in both domains. In the present day, the enomus challenges that educational institutions are facing the explosive growth of educational data and to use this data to enhance the quality of managerial decisions. Educational institutions are taking part in an important role in our society and playing a fundamental role for growth and development of nation. It is important to study and analyze educational data especially students’ performance.

To analyze the performance of the students, every institution has their own criteria. So, it is necessary to study on existing prediction techniques and hence to find the best prediction methodology for predicting the student academics progress and performance. A detailed literature survey on predicting student’s performance using data mining techniques has already been carried out in analyzing educational data. The main objective of this survey is to provide a great knowledge and examining different data mining techniques that have been applied to analyze and predict students’ performance.

Keywords - Educational Data Mining, Student Performance, Data Mining, Prediction, Classification Algorithm, Naïve Bayes algorithm.

I. INTRODUCTION

Education is a basic requirement to develop the country. The ability to change and to induce change and progress in society depends on education. The educational institutions are key resources for producing the good students to provide better services for the society. The main aim of higher education institutes is to give excellence education to its students and to improve the quality of managerial decisions [18]. Student Performance is measured how well the student has performed in class, and how well the student has understood the provided material. The quality of education delivered by educational institutes is measured by the performance achieved by the students. A quality education provides students with the skills they have to employ as productive, innovative and responsible individuals from society [6]. Education’s quality was evaluated by rate of success student’s and to what degree organization was able of preserved students. In education sector, the capability to predict performance of student was important [13].

Data mining is a computer-aided process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining techniques predict behaviors and future trends, allowing decision makers to make proactive, knowledge driven decision. Educational data mining is a recent research area that focuses on the use of data mining techniques to transform large volumes of educational data into useful and relevant knowledge that can improve the educational processes and decisions [5].

To provide quality education to learners, deep analysis of previous records of the learners can play a vital role. This paper aims to identify related factors and how predicting the performance of students can be done in educational domain by using different datamining methods. Predicting the student performance is the most essential tool to provide the
students and schools in improving their learning and teaching process.

The main objectives to this proposed work are:

i. To analyse the different usage of various prediction techniques of data mining in education

ii. To determine different student attributes which are mainly useful for the predicting the student performance

iii. To review the different prediction methods which are mainly effective for predicting the student’s performance.

II. COMPREHENSIVE REVIEW OF LITERATURE

This survey paper presents a detailed comprehensive literature review of major important researches in the area of Educational Data Mining for prediction of student performance. This survey will provide the teachers and leaners determine which datamining algorithm is suitable with existing students’ parameters and they can realize which parameters of the students are stored to predict their performance.

The referred work is presented below in a categorized tabular form Table 1. Table 1 show the complete survey based on evaluation of faculty performance of student dataset with various parameters using different data mining techniques in prior research studies from different perspectives.

This survey performed on the basis of the following categories:

(1) Objective of the research
(2) Parameters used for finding the faculty performance evaluation
(3) Methodology used in the research papers for evaluation of student performance
(4) Techniques and tools applied for evaluation
(5) Data Mining tools used for analysis and visualization
(6) Finding or conclusions of the research

<table>
<thead>
<tr>
<th>Sn.no.</th>
<th>Year &amp; Author</th>
<th>Paper Name</th>
<th>Objectives</th>
<th>Parameters Used</th>
<th>Methodology</th>
<th>Techniques and Tool Used</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>2017 A.K Hamoud, A.M. Humadi, W.A. Awadh, A.S. Hashim [1]</td>
<td>Students’ Success Prediction based on Bayes Algorithm</td>
<td>To analyze collective student information via a questionnaire and to classify the collected data to predict and categorize student performance</td>
<td>The questionnaire consists of 62 questions that cover the fields affecting students’ performance the most. The questions refer to health, social activity, relationships and academic performance.</td>
<td>The first stage is finding the most correlated questions to the final class, and the second is applying algorithms and finding the optimal algorithm.</td>
<td>Two built Bayes Algorithms (naïve Bayes and Bayes network) and Weka 3.8 were used.</td>
<td>The naïve Bayes algorithm is selected as an optimal choice for students’ success prediction.</td>
</tr>
<tr>
<td>2.</td>
<td>2015 F.Ahmad, Nur Hafieza Ismail, Azwa Abdul Aziz [12]</td>
<td>The Prediction of Students’ Academic Performance Using Classification Data Mining</td>
<td>To propose a framework for predicting students’ academic performance of first year bachelor students in Computer</td>
<td>8 year period intakes from July 2006/2007 until July 2013/2014 that contains the students’ demographics, previous academic records, and family background</td>
<td>Decision Tree, Naïve Bayes, and Rule Based classification techniques are applied to the students’ data in order to produce</td>
<td>Decision Tree, Naïve Bayes, and Rule Based</td>
<td>1. Rule Based is a best model among the other techniques. 2. The extracted knowledge from</td>
</tr>
<tr>
<td>Technique</td>
<td>Science course information</td>
<td>the best students’ academic performance prediction model.</td>
<td>prediction model will be used to identify and profile the student to determine the students’ level of success in the first semester.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>----------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applicatio of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study</td>
<td>To predict the performance of students based on their historical performance of grades by using the effectiveness of machine learning techniques</td>
<td>Student dataset consist of the academic records of 335 students. The total number of historical records of students’ grades was 6358. The periods analyzed were from the semester 2016-1 to the semester 2018-2 in the Computer Systems Engineering Degree of a university in Ecuador. In addition, the dataset comprises a total of 68 subjects organized into seven knowledge areas (Programming and Software Development, Mathematics and Physics, Information Network Infrastructure, Electronics, Databases, Economy—Administration, Machine learning tools, the decision tree algorithm, the visualization.</td>
<td>To establish the best correlations between the input variables and the result, which is the prediction of whether the student will pass a certain subject or not.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Authors</td>
<td>Title</td>
<td>Abstract</td>
<td>Keywords</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>A. Daud, N. Radi Aljohani, R. Ayaz Abbasi [2]</td>
<td>Predicting Student Performance using Advanced Learning Analytics</td>
<td>To predict whether a student will be able to complete his degree or not To investigate two new feature sets family expenditures and student personal information To construct an effective feature set of twenty-three features by combining proposed features along with exiting features.</td>
<td>General Education—Languages, Data Mining Techniques, Student Performance Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>2013 D Kabakchieva [8]</td>
<td>Predicting Student Performance by Using Data Mining</td>
<td>To find out if there are any patterns in the available data that could be useful for predicting</td>
<td>The dataset contains about 10330 students, described by 20 parameters, including gender, birth year, birth</td>
<td>Each classifier is applied for two testing options – cross validation</td>
<td>WEKA classifiers are used in the experimental</td>
<td>The decision tree classifier (J48) performs best (with the</td>
</tr>
</tbody>
</table>
Methods for Classification

students’ performance at the university based on their personal and pre-university characteristics

place, living place and country, type of previous education, profile and place of previous education, total score from previous education, university admittance year, admittance exam and achieved score, university specialty/direction, current semester, total university score, etc.

(using 10 folds and applying the algorithm 10 times – each time 9 of the folds are used for training and 1 fold is used for testing) and percentage split (2/3 of the dataset used for training and 1/3 – for testing). The WEKA J48 classification filter is applied on the dataset during the experimental study.

study, including a common decision tree algorithm C4.5 (J48), two Bayesian classifier (NaiveBayes and BayesNet), a Nearest Neighbour algorithm (IBk) and two rule learners (OneR and JRip).

highest overall accuracy), followed by the rule learner (JRip) and the k-NN classifier. The Bayes classifiers are less accurate than the others.


To solve the allocation problem of new students by using Bayesian classification.

The training data sets contain 50 students whose are described by attributes: end semester marks, class test grade, seminar performance, assignment, general proficiency, attendance, and lab work. The class label attribute and class type, have four distinct values namely first, second, third and fail.

Bayesian, the statistical classifiers, can predict class membership probabilities such as the probability that a given tuple belongs to a particular class. Bayesian classifiers assume an attribute value on a

Bayesian classifier

The performance of Bayesian classification technique is more suitable compared to rest of techniques such as genetic algorithm method.
<table>
<thead>
<tr>
<th>Application</th>
<th>Year</th>
<th>Authors</th>
<th>Methodology</th>
<th>Data Source</th>
<th>Model</th>
<th>Algorithm</th>
<th>Accuracy</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.</td>
<td>2014</td>
<td>Elakia, Gayathri, Aarthi, Naren J</td>
<td>Application of Data Mining in Educational Database for Predicting Behavioural Patterns of the Students</td>
<td>To suggest career options for the high school students and also to predict the potentially violent behaviour among the students by including extra parameters other than academic details</td>
<td>High school students based on every student’s interests, skills, likes, hobbies etc.</td>
<td>ID3 algorithm could be applied on them to learn the rules found on the training set and their corresponding labels.</td>
<td>ID3 algorithm outperforms every other decision trees. Hence ID3 algorithm was considered to be an applicable algorithm.</td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>2015</td>
<td>N. Puri, Deepali Khot, Pratiksha Shinde, Kishori Bhoite, Prof. Deepali Maste</td>
<td>Student Placement Prediction Using ID3 Algorithm</td>
<td>To identify relevant attributes based on academics, skills and curricular of final year student and design a model which can predict placement of the student using a classification technique based on decision tree</td>
<td>10th %, 12th % or Diploma, Degree Aggregate, Sem Performance, Communication Skill, Work on Projects, Internship, Education Gap, Backlog, Placement</td>
<td>Information Gain and entropy is calculated.</td>
<td>ID3 algorithm is best classifier with 95% accuracy</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td>S.Jayapra kash, Balamuru gan E.,Vibin Chandar</td>
<td>Predicting Students Academic Performance using Naive Bayes Algorithm</td>
<td>To predict student’s academic performance at the end semester exams by analyzing students feedback and their</td>
<td>The data was collected from 700 students in various departments of Blue Crest College, Accra, Ghana in the academic year 2014 with the internal examination score. The data are supervised learning approach through Naives Bayes algorithm</td>
<td>Naive Bayes algorithm</td>
<td>The comparison between feedback and internal examination marks. Naive Bayes algorithm gives the better</td>
<td></td>
</tr>
</tbody>
</table>
III. CONCLUSION

Data mining is an extremely vast area that includes applying different techniques and algorithms for pattern finding. Because of the data growth progresses, there were some difficulties for people to analyse and process large amount of data but with the help of data mining and machine learning techniques the analysis of such a huge amount of data has become easier. This paper has surveyed previous studies on predicting students performance with various analytical methods. According to the survey of applications on various Datamining methods, ID3 algorithm is the best predictor to predict the student performance. This survey showed how useful data mining can be applied in education particularly to improve students’ performance.

REFERENCES


[8] D. Kabakchieva, “Predicting Student Performance by Using Data Mining Methods for Classification”, ISSN: 1311-9702; Online ISSN: 1314-4081 ,DOI: 10.2478/cait-2013-0006


Applications of Data Mining Techniques in Student Performance Analysis A Survey


