PREDICTING THE REGISTRATION OF STUDENTS FOR AN EDUCATIONAL PROGRAMME USING DATA MINING TECHNIQUE

Prashant G. Tandale  
Dept. of Computer Applications  
Bharati Vidyapeeth Deemed University, Institute of Management  
Kolhapur, Maharashtra

Abstract— Educational data mining is an emerging discipline that focuses on applying data mining tools and techniques to educationally related data. The data mining technology can discover the hidden patterns, associations, and anomalies from educational data. This knowledge can improve the decision making processes in higher educational systems. Data mining is considered as the most suited technology appropriate in giving additional insight into the lecturer, student, alumni, manager, and other educational staff behavior and acting as an active automated assistant in helping them for making better decisions on their educational activities. This paper tries to take review of work done so far in this area of educational data mining and one of the applications of data mining called “Prediction”.

Keywords— KDD, Educational Data Mining (EDM), Classification, WEKA, Decision Tree, Prediction.

I. INTRODUCTION

Education enables the people of a country civilized and well mannered. To remain competitiveness among educational field, these organizations need deep and enough knowledge for a better assessment, evaluation, planning, and decision-making. The required knowledge cannot be gained from the tailor made software used now a day. Data mining incorporates a multitude of techniques from a variety of fields including databases, statistics, data visualization, machine learning and others. Educational data mining (EDM) is an emerging discipline that focuses on applying data mining tools and techniques to educationally related data. The discipline focuses on analyzing educational data to develop models for improving learning experiences and improving institutional effectiveness [14]. Discovering hidden patterns, associations, and anomalies from educational data can be done using data mining technology. This knowledge can improve the decision making processes in higher educational systems. Data mining is considered as the most suited technology appropriate in giving additional insight into the all stake holders of the educational institute. The data mining technology can discover hidden patterns, associations, and anomalies from educational data can be done using data mining technology. This knowledge can improve the decision making processes in higher educational systems. Data mining is considered as the most suited technology appropriate in giving additional insight into the all stake holders of the educational institute.

II. REVIEW OF LITERATURE

A literature review on educational data mining topics such as student retention and attrition, personal recommender systems within education, and how data mining can be used to analyze course management system data.

According to Paulraj and Ponniah, the main benefits of data mining to educational institutes are – It provides an integrated and total view of an institute. It makes the institute’s current and historical information easily available for the decision making. It provides the facility to students to get their different subject notes from a web enabled database. It provides the information about student’s attendance. Students can get theirs results easily and very quickly. It helps to provide information about faculty like how many members are there in all the different departments etc [6].

Minaei-Bidgoli & others proposed an approach to classify students in order to predict their final year grade based on the features extracted from logged data in an educational web based system was reported. Data mining classification process was used in conjunction with genetic algorithm to improve the prediction accuracy.

C. Romero and S. Ventura carried a survey in the field of education. They have described the types of users, types of educational environments and the data they provide. Also they have explained in their work the common tasks in the educational environment that have been resolved through data mining techniques [9].

Hua-long Zhao has done Multidimensional cube analysis by taking use of OLAP technology and has shown that the curriculum chosen by the students can depend upon many angles like teacher, semester and student. He has used Star model of data warehouse to the analysis of curriculum which can provide certain policy making support for different levels of education policy-maker in the school.

Fadzilah Siraj and Mansour Ali Abdoulha have used data mining techniques for understanding student enrolment data. They have done comparative study of three predictive data mining techniques namely Neural Network, Logistic...
regression and Decision tree. The results obtained can be used by the planners to formulate proper plan for the university [12].

Shaeela Ayesha & others discusses data mining technique named k-means clustering is applied to analyze student’s learning behavior. Here K-means clustering method is used to discover knowledge that come from educational environment [11].

S K Altaf Hussain Basha have identified the association between different attributes of educational environment i.e. the location of the college, type of college, different social groups, different courses etc. and thereby extract strong association rules. They used data mining technique of association rule mining to extract strong rules in educational environment that identifies students’ success patterns in different colleges in different social groups. Further they have processed the available data to find the pattern of support for these rules from time to time [1].

Ramasubramanian P. states that Rough Set Theory (RST) has found many interesting applications in academia. Currently there is an increasing interest in data mining and educational systems, making educational data mining as a new growing and promising area for the research community. Decision making of classroom processes is a very tedious process which involves observing a student’s behavior, analyzing historical data, and estimating the effectiveness of pedagogical strategies. RST has been used in estimating the teaching result analysis with the help of concept map and brainstorming which can be used as a tool for educators [7].

Robertas analyzed student academic results for informatics course improvement, rank course topics following their importance for final course marks based on the strength of the association rules and proposed which course specific course topic should be improved to achieve higher student learning effectiveness and progress [8].

W.M.R. Tissera & others presents a real-world experiment conducted in an ICT educational institute in Sri Lanka. A series of data mining tasks are applied to find relationships between subjects in the undergraduate syllabi. This knowledge provides many insights into the syllabi of different educational programmes and results in knowledge critical in decision making that directly affects he quality of the educational programmes [13].

Hongjie Sun conducts a research on student learning result based on data mining. It is aimed at putting forward a rule-discovery approach suitable for the student learning result evaluation and applying it into practice so as to improve learning evaluation skills and finally better serve learning practicing [4].

S. Anupama Kumar and Dr. Vijayalakshmi M.N applied decision tree algorithm on student’s internal assessment data to predict their performance in the final exam. The outcome of the decision tree predicted the number of students who are likely to fail or pass [10].

Monika Goyal used different types of rule-based systems and have been applied to predict student’s performance (mark prediction) in an e learning environment (using fuzzy association rules). Several regression techniques are used to predict student’s marks like linear regression for predicting student’s academic performance, stepwise linear regression for predicting time to be spent on a learning page, multiple linear regression for identifying variables that could predict success in courses and for predicting exam results in distance education courses [2].

According to Jaiwei Han, data mining is an interdisciplinary field of astronomy, business, computer science, economics and others to discover new patterns from large data sets. The actual data mining task is to analyze large quantities of data in order to extract previously unknown patterns such as groups of data records (cluster analysis), unusual records( anomaly detection) and dependencies (association rule mining) [3].

III. KNOWLEDGE DISCOVERY PROCESS

Data mining is also known as Knowledge Discovery from Databases (KDD). Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. At the level of a university, administrative information systems collect information about students, their enrolment in particular programs and courses, and performance like examination grades. In addition, the information about the lectures, instructors, study programs, courses and prerequisites, are typically available as well. These data can be analyzed from various levels and perspectives, showing different aspects of organization, and giving us more insight in the overall educational system which is shown in Figure1. [16]
IV. Techniques Used in Data Mining

A. Association analysis

Association analysis is the discovery of association rules showing attribute-value conditions that occur frequently together in a given set of data. Association analysis is widely used for market basket or transaction data analysis.

B. Classification and Prediction

Classification is the processing of finding a set of models (or functions) which describe and distinguish data classes or concepts, for the purposes of being able to use the model to predict the class of objects whose class label is unknown. The derived model may be represented in various forms, such as classification (IF-THEN) rules, decision trees, mathematical formulae, or neural networks. Classification can be used for predicting the class label of data objects. However, in many applications, one may like to predict some missing or unavailable data values rather than class labels. This is usually the case when the predicted values are numerical data, and is often specifically referred to as prediction. IF-THEN rules are specified as IF condition THEN conclusion e.g. IF age=youth and gender=female then buys_dress=yes

C. Clustering Analysis

Unlike classification and prediction, which analyze class labeled data objects, clustering analyzes data objects without consulting a known class label. In general, the class labels are not present in the training data simply because they are not known to begin with. Clustering can be used to generate such labels. The objects are clustered or grouped based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity. That is, clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Each cluster that is formed can be viewed as a class of objects, from which rules can be derived.

D. Outlier Analysis

A database may contain data objects that do not comply with the general behavior of the data and are called outliers. The analysis of these outliers may help in fraud detection and predicting abnormal values.

V. Applications of Data Mining Techniques in Higher Education

A. Predicting The Registration of Students in an Educational Programme

Educational organizations are getting strong competition from other Academic competitors. To have an edge over other organizations, needs deep and enough knowledge for a better assessment, evaluation, planning, and decision making. Data Mining helps organizations to identify the hidden patterns in databases; the extracted patterns are then used to build data mining models, and hence can be used to predict performance and behavior with high accuracy. As a result of this, universities are able to allocate resources more efficiently. One of the applications of data mining can be for example, to efficiently assign resources with an accurate estimate of how many male or female will register in a particular program by using the Prediction techniques. For this purpose Admissions to first year of UG courses data are collected from the BDVU, Institute of Management, Kolhapur and it is analyzed using WEKA- Data mining Tool and weka.classifiers.rules.ZeroR scheme.

The are two attributes taken into consideration are Academic Year and Number of Admissions. By using classifier model that is full training set ZeroR predicts the class value as 47. That means the expected number of students in academic year 2015-16 is 47.

B. Predictor Training Data:

<table>
<thead>
<tr>
<th>Year of Admission</th>
<th>No. of students Admitted to BCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-11</td>
<td>53</td>
</tr>
<tr>
<td>2011-12</td>
<td>47</td>
</tr>
<tr>
<td>2012-13</td>
<td>23</td>
</tr>
<tr>
<td>2013-14</td>
<td>56</td>
</tr>
<tr>
<td>2014-15</td>
<td>56</td>
</tr>
<tr>
<td>2015-16</td>
<td>?</td>
</tr>
</tbody>
</table>

Predicted Student Count= 47

VI. Conclusion

In this paper we have taken a review of the work so far done in applying data mining techniques in higher education so that it can generate strategic information which will help the higher authorities in decision making. Since the application of data mining brings a lot of advantages in higher learning institution, it is recommended to apply these techniques in the areas like optimization of resources, prediction of admissions, etc. We can plan according for the next academic year. Also we could take necessary actions to increase the admission count.

Acknowledgment

Dr. A T Gaikwad special thanks for motivating and guiding me to write this paper. Dr. Nitin Nayak my mentor, always the source of positive energy and motivator.

References

...

Fig.2 Result of ZeroR Test using WEKA tool