

Recommendation of Data Mining Technique in Higher Education

Prof. Priya Thakare Guide

Ajinkya Kunjir, Poonam Pardeshi, Shrinik Doshi, Karan Naik

Department of Computer Engineering Sinhgad Institute of Technology and Science Pune, India

ABSTRACT

In this paper we will discuss about the problem that are faced by higher education institutions. One of the biggest challenges that higher education faces today is predicting the right path of students. Institutions would like to know, which students will enroll in which course, and which students will need more assistance in particular subject and what efforts should be taken for weak students. Also some time management needs more information about student like their overall result, interest in co-curricular and extra-curricular activities and about the success of new offered courses. One way to effectively address the challenges for improving the quality of students and education is to provide new knowledge related to the educational processes and entities to the system. This knowledge can be extracted from historical data that reside in the educational organization's databases using the techniques of data mining technology. If data mining techniques such as clustering, decision tree, association, classification and prediction can be applied to higher education processes, it can definitely help improve students' overall performance, their life cycle management, selection of course and predict their dropout rate.

KEYWORDS: Data mining, Higher education, Clustering, Decision tree, neural network, classification, prediction, association rule analysis.

I. INTRODUCTION

One of the significant facts in higher learning institution is the explosive increase of educational data. These data are increasing rapidly without any benefit to the management and institutions. We believe that to manage this vast data is difficult task, but by new techniques and tools we can easily process the large amount of generated data in business processes and extract some useful knowledge and information from it. Data mining is a technique of extraction hidden predictive information from large databases; it is a powerful new technology with great potential to help higher learning Universities or institutions to focus on the most important information in their data warehouses. Data mining tools predict future trends and behavior patterns, allowing institution to make proactive, knowledge-driven and appropriate decisions. The automated, prospective analyses offered by data mining technology move beyond the analyses of past events provided by retrospective tools typically of decision support systems. Data mining tools can answer institution questions that traditionally were too time consuming in past to resolve. Higher education institutions can use classification technique, for a comprehensive analysis of student characteristics, or use estimation and prediction technique to predict the likelihood of a variety of outcomes, such as transferability, choosing elective, choosing right career path, drop out and course success.

A. Data mining tools and algorithms

Machine Learning
Artificial Intelligence
Emulating human intelligence
Neural Networks for prediction
Biological models and psychological models.
SLIQ (Supervised Learning in Quest)

B. Phases of Data Mining

Data mining is an iterative process that typically involves the following phases:
Problem definition
Data exploration
Data preparation
Modeling
Evaluation

C. Tools of Data Collection & Analysis

Various tools are needed for project are for analyzing data, designing, implementation and some developing software tool such as:

- MYSQL DATABASE
- EXCEL
- MS ACCESS
- SPSS
- METLAB TOOL
- WEKA DATA MINING TOOL
- TANGARA DATA MINING TOOL
- WEB MINER
- V.B 6.0

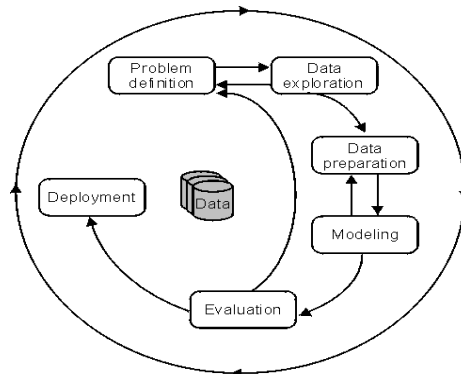


Fig 1.1 Data Mining Process

II. II.DATA MINING: A WAY TO IMPROVE TODAYS HIGHER LEARNING INSTITUTIONS

Data mining is a powerful, new and emerging technology with great potential in information system. It can be best defined as the automated process of extracting useful knowledge and information including, patterns, associations, trees, changes, trends, anomalies and significant structures from large or complex data sets that are not classified. Our main idea is that the hidden patterns, associations, classification and anomalies that are discovered by data mining techniques can help bridge this knowledge gap in higher learning institutions. The knowledge discovered by data mining techniques would enable the higher learning institutions in making better decisions, having more advanced planning in directing students, predicting individual behaviors with higher accuracy, predicting the dropout rates and enabling the institution to allocate resources and staff more effectively. It results in improving the quality, effectiveness and efficiency of the processes. The term data mining is often used to apply to the two separate processes of knowledge discovery and prediction. Knowledge discovery provides explicit information that has a readable form and can be understood by a person at user end. Forecasting, or predictive modeling provides predictions of future events which can help in betterment and may be transparent and readable in some approaches and opaque in others such as neural networks. Data mining relies on the use of real world data. These data are extremely vulnerable to co-linearity because data from the real world may have unknown relations with each other. Data mining is the entire process of applying computer-based methodology, including new techniques and technologies for knowledge discovery. This paper presents how various data mining techniques can be implemented in the field of higher education to discover some meaningful patterns or relations that can further improve the overall performance and quality of higher education and students' respectively.

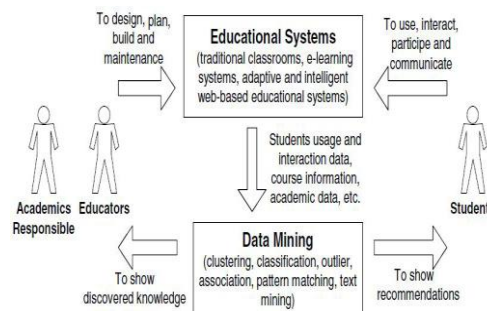


Fig 2.1. The cycle of applying data mining in educational systems

III. III.PROPOSED ANALYSIS GUIDELINE (DM-HEDU) FOR APPLICATION OF DATA MINING IN HIGHER LEARNING INSTITUTION

In this section, we propose a new analysis guideline to present a roadmap or the area of data mining application in higher learning institution. Its adopted and primary name is DM-HEDU (Data Mining in Higher Education System). As today’s higher learning institutions deal with powerful and strong business competitors in a highly competitive environment, they have to look for a new, faster and innovative solution to overcome the problems and achieve a high academic institutional standard. Therefore, this guideline may assist the institutions and organizations to identify the ways to improve their processes and help to take decisions. In the previous literature studies we have not discovered a complete guideline which gathers most of the possible processes to improve the level of higher education learning institution through data mining. The idea of the proposed guideline is presented in Fig. 1.1 The importance of DM-HEDU guideline in a higher education learning institutions can be viewed from different perspectives as follows:

Our DM-UEDU guideline can be used for unifying a common context to identify the current gaps and further works in future for any data mining application in a higher education based on the processes of higher education learning institution. It also provides a great opportunity for researchers to be known with the existing area of research and development in the field of higher education learning. There are numerous areas in which data mining can be applied. Following table shows the attributes which can be used in higher education learning:

ID	Objective	Data Mining Method	Explicit Knowledge	Educational Process
1	Use of Data Mining in correct scores	Prediction	The patterns of previous student test score associated with their marks, attendance ,extra-curricular activities and so on.	Planning-Course assessment.
2	Creating meaningful learning outcome typologies	Cluster analysis	The patterns are generated based on previous student’s learning outcome	Evaluation-Student assessment
3	Academic planning and intervention transfer prediction	Prediction	The success rate patterns of previous students who had previously transferred subjects	Evaluation-Student assessment
4	Predicting students’ overall performance	Classification	Classified pattern of previous students based on their performance throughout the year	Evaluation-Student assessment
5	Improving quality of graduate student by data mining approach	Association, Classification	Characteristic patterns of previous students who took a particular major and The patterns of previous students which were likely to be good in a given major Counseling	Student major counseling

Table 3.1: Summary analysis of previous study

A.ID 1: Uses of Data Mining in CRCT Scores

This study (Gabrilson, 2003) attempts to analyze the most effective factor in determining students’ score in various subjects. It presents that the useful and meaningful discovered patterns targeting the various relationship of different types of variables are the major factors affecting the students test score. Using data mining prediction technique, these factors are thus successfully identified.

From this case study the effectiveness of data mining in predicting the most effective and necessary factors in student test score can be concluded. It results in improving the evaluation and also helps in student assessment process. Improving this process has a direct impact in improving transition rate of a higher learning institution and thus decreases the drop-out rate.

B.ID 2: Creating Meaningful Learning Outcome Typologies

This case study aims at creating meaningful learning outcome typologies using data mining techniques. The main objective of obtaining typologies of students is to be able to improve students' performance through predefined clusters of behavior. These clusters help higher education universities to better identify the requirements of each group and make better decision on how to behave with them in terms of educating, offering courses and curriculum, required time for teaching and so on. It results in having more student satisfaction of their studies, course offering, and class's lectures. From this study we can conclude the effectiveness of data mining in developing typologies of students in higher educational domain. The result has a great impact in improving educational achievement of a higher education learning through improving the evaluation-student assessment process.

C.ID 3: Academic Planning and Interventions Transfer Prediction

This case study presents data mining advantages in predicting students' likelihood of transferability for on time intervention. It notifies the institution the types of students who are most at risk of not transferring to a higher level before they know about it. The outcome enables institutions in predicting the likelihood of student transferability. Data mining can link student's academic behavior with their final transfer outcomes. Therefore these kinds of identifications help the institutions to pay more attention to those who require more academic assistance by setting extra classes, consultation hours with the university's counselors and psychologists.

From this study, we conclude the predictions of student's likelihood of transferability assist decision makers with an additional tool to identify those who are not much likely to transfer. As a result, this prediction provides a great potential impact on improving the transition rate of an educational university through improving the student assessment process of an educational domain.

D.ID 4: Predicting a Student's Performance

This case study uses the data mining classification technique to predict the students final grades based on their web-use feature. By discovering the successful patterns of students in various categories, the institute can predict the final grade of each single student. Therefore it helps to identify students who are at risk early and allow the faculty to provide appropriate advice.

From this case study, we can conclude that data mining is effective in predicting student's performances in the educational domain. The result has deep impact in improving the transition rate, and the process indicator of a higher learning institute by improving the student assessment process to some extent.

E.ID 5: Improving Quality of Graduate Students by Data Mining

A study is done by Kitsana (2003) to improve the quality of graduate students with the help of data mining. The objective of our project is to discover and study the knowledge from large sets of engineering student's databases records. The discovered knowledge in the form of patterns is useful in assisting the development of new curricula, improving of existing curricula and most important, helping students to select the appropriate elective. The final result represents the most appropriate elective for each single student. The extracted patterns are useful for university counselors or supervisors who are supposed to counsel and supervise new students.

F. Advantage of Data mining in Academics

Data mining gives the answers of questions like:

- Q: Who is the weak/strong student?
- Q: Who are the students taking most credit hours?
- Q: Which is the interesting subject of the students?
- Q: What type of course can institute offer to attract more students?
- Q: How can faculty help weak students?
- Q: How the overall college result can be improved?
- Q: Is the teaching pattern satisfactory or need to be changed?
- Q: Which is the most appropriate elective for the student?

IV. DATA ANALYSIS AND INVESTIGATION

A. Domain Understanding

In this phase the higher education is wholly analyzed and the main data mining objectives are set and targeted accordingly.

B. Data Understanding

In this phase, the required raw data and attributes of students and faculties are collected based on the pre-defined objectives. According to our data mining goal, the raw data is related to:

- 1) Student demographic and academic knowledge.
- 2) Lecturer demographic and academic knowledge.

- 3) Course information and contents.
- 4) Semester status information and planning.

The data are then described and explored by (i) identifying pre-defined initial format of data, (ii) the meaning and description of individual attributes of student and faculty and (iii) determining the relation of attributes with each other. The final part verifies the quality of data by determining the completeness and correctness of data.

C. Data Preparation

This phase of data mining is the final step of directly dealing with data. The dataset produced in this section is used for modeling and the major analysis task. The importance of data preparation is to maximize visibility of the relationship that exists between input and output data sets, which is captured with a modeling tool. Prepared data enables data mining technique to generate a better and efficient model.

V. DATA MINING MODELING

The knowledge obtained from data mining techniques gives the managerial decision makers the useful information for taking proper decisions. The models are classified in two main categories: predictive models and descriptive models.

- Descriptive model describes the data set in a concise and summarized manner and presents the interesting and important general properties of the data. It explains the extracted patterns in existing data, which may be used to guide managerial decisions.
- Predictive model predicts behavior based on previous data and uses data with known results to build a model that can be used in future to explicitly predict values for different data (Two Crows Corporation, 1999).

A. Predictive Data Mining Models

Model A: Predicting Student Success Rate for Individual Student

This model is developed to predict the student success rate for individual students and to predict the student dropout rate. The explicit knowledge discovered from this model can be used by student management system to consult individual student based on his performance in successful course taking and choosing appropriate electives. Within this procedure, if the students are predicted to be unsuccessful, then they are provided with extra consultation and extra efforts are taken by faculty and the universities to help them to get improved in the course. If they are successful, new personal policies for the successful student course taking are set.

For classification method, the experiments are conducted through decision tree using Supervised Learning In Quest algorithm (SLIQ) (Mehta et al., 1996) and neural classification. For prediction method the experiments are conducted using neural prediction (Han and Kamber, 2001) and Radial Basis Function (RBF).

Model B: Predicting Student Success Rate for Individual Lecturer

This model is developed to predict the student success and failure rate for individual lecturer. The explicit knowledge discovered from this model can be used by management for general and managerial decision-making. It can be used to support policies and procedures, which are set at top-level management. This model is applied through prediction (neural network) and classification (SLIQ and neural network) techniques.

Model C: Model of Lecturer Course Assignment Policy Making

This model is developed to describe the characteristics pattern and way of teaching of lecturers who plan to take the course. The knowledge discovered from this model can be used for general decision making at top-level management. It helps in knowing how the lecturer teaches. i.e. is his teaching pattern fruitful or not. It assists in supporting the current managerial rules and regulations in lecturer course assignment policy making and it also helps to set new strategies and plan for managerial decision makers on those lecturers who plan to conduct the course.

V. CONCLUSION AND FUTURE WORK

The current education system does not involve any prediction about fail or pass percentage of students based on their performance. The system doesn't deal with student dropouts. Since the proposed model identifies the weak and lagging students, the teachers can provide support and academic help for them. It also helps the teachers to act before a student drops or plan for recourse allocation with confidence gained from knowing how many students are likely to pass or fail. Among several innovation in recent technology, data mining is making a great impact and comprehensive changes in the field of higher education. Such activities will definitely guide to better decision making procedures and will improve the quality of education.

As a further work, we would like to enhance other data mining processes in higher learning institution by referring to DM-HEDU analysis guideline. These processes are according to first class priorities of the institutions. Other work can be generating student and lecturer models for the other type of course offered in the

institute. Since the application of data mining brings a lot of advantages in higher learning institution and even helps in improving the quality of students and education, it is recommended to apply these techniques in other academic institution like primary and secondary schools, language institutions, institutions for special students and private and government colleges especially in India.

ACKNOWLEDGEMENT

This work is carried under precious guidance of Prof. Priya Thakare, Dept. of Computer Engineering, SITS, Narhe. The authors are thankful for such helpful guidance and discussions at the early stage of the project.

REFERENCES

- [1] Agrawal, R.T. and Imielinski, A.S. (1993). Mining association rule between sets of item in large database. In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D. C., 207–216.
- [2] Cave, M., Kogan, M. and Hanney, S. (1990). The scope and effects of performance measurement in British higher education. In F.J.R.C. Dochy, M.S.R. Segers and W.H.F.W. Wijnen (Eds.), Management Information and Performance Indicators in Higher Education. Van Gorcum and Comp, B.V., Assen/Maastricht, 48–49.
- [3] Chang, W.H.T. and Lee, Y.H. (2000). Telecommunications data mining for target marketing. *Journal of Computers*, 12(4), 60–74.
- [4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000). CRISP-DM1.0: Step-by-Step Data Mining Guide.
- [5] Chen, M.S., Han, J. and Yu, P.S. (1996). Data mining: an overview from a database perspective. *IEEE Transaction on Knowledge and Data Engineering*.
- [6] Delavari, N. and Beikzadeh, M.R. (2004). A new analysis model for data mining processes in higher educational systems. In MMU International Symposium on Information and Communications Technologies 2004 in Conjunction with the 5th National Conference on Telecommunication Technology 2004. Putrajaya, Malaysia
- [7] Delavari, N., Beikzadeh, M.R. and Shirazi, M.R.A. (2004). A new model for using data mining in higher educational system. In 5th International Conference on Information Technology Based Higher Education and Training: ITEHT '04. Istanbul, Turkey.
- [8] Delavari, N., Beikzadeh, M.R. and Phon-Amnuaisuk, S. (2005). Application of enhanced analysis model for data mining processes in higher educational system. In 6th International Conference on Information Technology Based Higher Education and Training. Santo Domingo, Dominican Republic.
- [9] Edelstein, H. (2000). Building profitable customer relationships with data mining. SPSS White Paper-Executive Briefing. Two Crows Corporation.