Data Mining Applications in Higher Education

A case Study of BVU IM, Kolhapur

Dr. K M Alaskar
Professor, Dept. of Computer Applications
Bharati Vidyapeeth University, Inst. of Management
Kolhapur, Maharashtra, India
k.alaskar@rediffmail.com

Prof. Prashant G. Tandale
Asst. Professor, Dept. of Computer Applications
Bharati Vidyapeeth University, Inst. of Management
Kolhapur, Maharashtra, India
prashant_gs@yahoo.com

Prof. A A Basade
Asst. Professor, Dept. of Management Studies
Bharati Vidyapeeth University, Inst. of Management
Kolhapur, Maharashtra, India

Abstract— *One of the biggest challenges that higher learning institutions face today is to improve the quality of managerial decisions. The managerial decision making process becomes more complex as the complexity of educational entities increase. Educational institute seeks more efficient technology to better manage and support decision making procedures or assist them to set new strategies and plan for a better management of the current processes. One way to effectively address the challenges for improving the quality is to provide new knowledge related to the educational processes and entities to the managerial system. This knowledge can be extracted from historical and operational data that reside in the educational organization's databases using the techniques of data mining technology. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from large data sets. This paper is based on the use of data mining to analyze the students' feedback on curriculum.*

Keywords—Data Mining, Curriculum, Student's Feedback Database, Average Rating, Performance indicators

INTRODUCTION

One of the biggest challenges that higher education faces today is predicting the paths of students. Institutions would like to know, for example, which students will enroll in particular course programs, and which students will need assistance in order to graduate. Are some students more likely to transfer than others? In addition to this challenge, traditional issues such as enrollment management and time-to-degree continue to motivate higher education institutions to search for better solutions. One way to effectively address these student and alumni challenges is through the analysis and presentation of data, or data mining. Data mining enables organizations to use their current reporting capabilities to uncover and understand hidden patterns in vast databases. These patterns are then built into data mining models and used to predict individual behavior with high accuracy. As a result of this insight, institutions are able to allocate resources and staff more effectively. Data mining may, for example, give an institution the information necessary to take action before a student drops out, or to efficiently allocate resources with an accurate estimate of how many students will take a particular course. This white paper addresses the capabilities of data mining and its applications in higher education. Three case studies demonstrate how data mining can be used to study the opinion of students about syllabus.

DATA MINING OVERVIEW

Data mining uses a combination of an explicit knowledge base, sophisticated analytical skills, and domain knowledge to uncover hidden trends and patterns. These trends and patterns form the basis of predictive models that enable analysts to produce new observations from existing data. Gartner Inc.'s definition of data mining is the most comprehensive: "The process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories, and by using pattern recognition technologies, as well as statistical and mathematical techniques." Data mining should be performed on very large or raw datasets using either supervised or unsupervised data mining algorithms. Note that data mining cannot occur without direct interaction with unitary data. Data mining is an evolving field with new concepts born monthly and current concepts struggling to retain their place. Many of the new and interdisciplinary concepts, such as the stochastic search methods (including genetic algorithms), market basket analysis, memory based reasoning, and Bayesian averages, were not even imagined less than a decade ago. Researchers from different branches of mathematics, statistics, marketing, or artificial intelligence will use different terminologies. Where a statistician sees dependent and independent variables, and an artificial intelligence researcher sees features and attributes, others see records and fields (Berry

and Linoff, 1997). The phrase "neural networks" is synonymous with data mining. Although data mining is known for having exotic names, the field has begun to include certain kinds of descriptive statistics and visualization techniques into data mining (Westphal and Blaxton, 1998). Statsoft, an on-line statistical software provider, seemed to favor "exploratory data analysis." Berthold and Hand (1999) called their work "intelligent data analysis." Finally, in statistical language, Statsoft (2001) categorizes typical on-line analytical processing (OLAP) techniques as basic statistical exploratory methods or exploratory data analysis that include such techniques as "examining distributions of variables (e.g., to identify highly skewed or non-normal, such as bi-modal patterns), reviewing large correlation matrices for coefficients that meet certain thresholds, or examining multi-way frequency tables." It reserves the term "multivariate exploratory techniques" for data mining. These techniques are designed specifically to identify patterns in multivariate (or univariate, such as sequences of measurements) data sets that include cluster analysis, factor analysis, discriminant function analysis, multidimensional scaling, log-linear analysis, canonical correlation, stepwise linear and nonlinear regression, correspondence analysis, time series analysis, and classification trees, etc.

BACKGROUND OF EDUCATIONAL SYSTEMS BASED ON INDICATORS

The Indicators are agreed measurement scales which identify the quantitative relationships between two variables. They are normally used as numerical values. Indicators are very important in determining the goals and the operational analysis of the educational system (Johnstone, 1976; Johnstone, 1981; Wako, 1988). The higher learning institution of a country deals with human factors and educating specialists needed by the community, educational promotion, research development, and providing a suitable environment for the country's growth. Thus, the system essentially requires a principle which can express the qualitative characteristics of the higher learning institution to some quantitative values, and facilitates evaluating the functionalities. This principle is summarized into indicators. To evaluate the different aspects of the higher learning institution, "performance indicator" is used as one of the main educational system indicators (UNESCO, 2006b). The earlier studies state that other than performance indicator, an additional step for supporting educational system improvement, which is built on information from performance indicator, is more important. This step is called educational feedback. The feedback should be up-to-date, valid and reliable. From the above literature survey in higher learning institution it is possible to derive the following conclusions:

1. Based on the fact that the performance feedback perceived in an educational institution should be accurate, up-to-date, reliable, valid, and toward the goals of educational improvements, therefore more effective strategies should be taken into account to improve the feedback from an educational domain. Not only the performance indicator is essential for indicating the actual state of an education system, it is also vital to develop a methodology for educational system performance feedback.
2. Improving the feedback of an educational domain implies further analysis and investigation in the forming components of performance indicator. Data mining is able to improve the educational system in each component of the performance indicator. This improves the feedback from the system.

In this study, the components of performance indicator are based on student's feedback on Curriculum. In India National Assessment And Accreditation council, Bangalore (NACC) has give some standard performance indicators for evaluating the curriculum by students. According to that feedback of 50 students out of 500 students has been taken using simple random sampling method. Here sample size is 50 i.e 10% of overall population of the study.

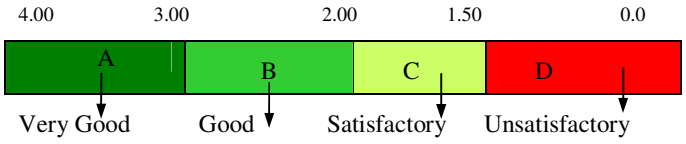
PERFORMANCE INDICATORS

Following table show the performance indicators used to analyze the feedback by students towards the curriculum.

PERFORMANCE INDICATORS

Parameter	***Meaning***
1	Depth of course content including project work if any
2	Extent of coverage of course
3	Applicability / relevance to the real life situations
4	Learning Values
5	Clarity and relevance of textual reading material
6	Relevance of additional source material(Library)
7	Extent of effort required by students
8	Overall Rating

4 POINT SCALE



DATA COLLECTION AND PRESENTATION

PARAMETER AND RATING OBTAINED

Para meter	A	B	C	D	Total	%
1	34	5	7	4	50	**12.47**
2	23	21	4	3	51	**12.72**
3	13	20	13	5	51	**12.72**
4	25	13	5	7	50	**12.47**
5	15	15	12	7	49	**12.22**
6	18	10	19	6	53	**13.22**
7	12	11	20	6	49	**12.22**
8	16	12	10	10	48	**11.97**
Total	156	107	90	48	**401**	
%	38.90	26.68	22.44	11.97		

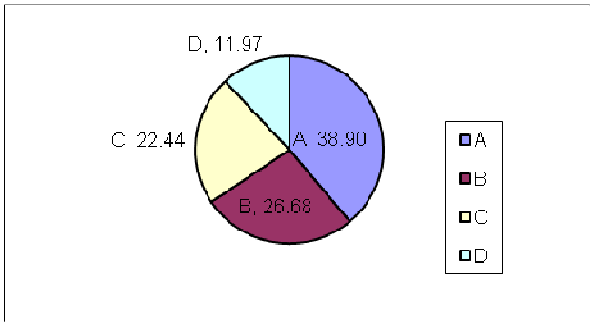


Fig. 1 Pie-chart showing overall rating obtained

From above pie chart we can see that maximum choice is for option A i.e 38.90 percentage. Option B and C has nearly same percentage i.e 26.68 and 22.44 respectively. Option D has the lowest choice i.e 11.97 percentage. The overall rating can be calculated as follow.

Overall Rating = [(Total of A * 4) + (Total of B * 3) + (Total of C *2) + (Total of D * 1)] / Grand Total

Overall Rating = [(156 * 4) + (107*3) + (90* 2) + (48 * 1)] / 401

Overall Rating = 2.93

This indicates feedback remark on curriculum by students is lies between 3 and 2 i. e 2.93 means Good.

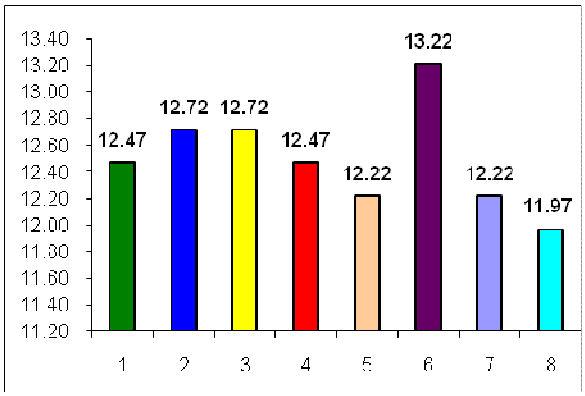


Fig. 2 Parameter wise percentage obtained

From above figure we can see that maximum percentage of students say that Relevance of additional source material from Library is required in the syllabus.

CONCLUSION AND FUTURE WORK

The result of this study indicates that Data Mining Techniques (DMT) capabilities provided effective improving tools for student feedback analusis. It showed how useful data mining can be in higher education in particularly to predict acceptance and changes of curriculum by students. We collected the data from student by using questionnaire to find the relationships between behavioral factors of student. For future work, application of data mining techniques in educational field can be used to develop performance monitoring and evaluation tools system. DMT has a potential in performance monitoring of High school and other levels education offering historical perspectives of students' performances. The results may both complement and supplement tertiary education performance monitoring and assessment implementations

REFERENCES

Jing Luan, "Data Mining Applications in Higher Education"

Naeimeh Delavari, Mohammad Reza, "Data Mining Applications in Higher Education" Information in Education 2008, Vol. 7 No. 1, pp 31-54

Han,J. and Kamber,M. "Data mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann series in Data Management System, Jim Grey, series Editor. 2006.

Berry, M., and Linoff, G. Master Data Mining: The Art and Science of Customer Relationship Management. New York: Wiley Computer Publishing, 2000

Sajadin Sembering, M Zarlis, Dedy Hartama " Prediction of Student Academic Performance By an Application of Data Mining Techniques , 2011 International Conference on Management and Artificial Intelligence, Vol. 6 ,2011

www.nacc-india.org