

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Conceptual Framework of Data Mining Process in Management Education in India: An Institutional Perspective

Jayanthi Ranjan and Saani Khalil
Information Management Area, Institute of Management Technology,
Raj Nagar, Ghaziabad, Uttar Pradesh, India

Abstract: The study introduces data mining in the context of management education. The role of management education offers great opportunity for many interesting and challenging data mining applications. The meaningful knowledge and potentially useful patterns extracted through data mining can assist in improving the quality of education and performance of students. We describe the conceptual frame work of data mining process in management education. Management institutions in India will find larger and wider applications for data mining as these institutes carry research and teachings that relates to creation, transformation and utilization of knowledge. The framework helps management institutes to explore the effects of probable changes in recruitments, admissions and courses and ensures efficiency in the quality of students, student assessments, evaluations and allocations. This study describes the data mining process in management education in general and academic aspects of admission and counseling process in particular. The patterns that can be generated using data mining techniques are also suggested. The study concludes with future directions, lessons learned and limitations of study.

Key words: Management education, students, faculty, courses, data mining, patterns

INTRODUCTION

Data mining has proven to be a powerful tool capable of handling decision making and forecasting techniques. The predictive power of data mining comes from machine learning, pattern recognition and statistics to automatically extract concepts and to determine interrelations and patterns of interest from large databases. It is our intent to explore how data mining is being used in industries and corporates and connect those techniques in management education.

Management institutions generate data about students, courses, faculty, staff that includes managerial systems, organizational personnel, lectures details and so on. This data which serves as a strategic input is very useful to any Management Institution for improving the quality of education process. Often there are several patterns we come across in evaluations, courses, students' counseling and admissions. These patterns if extracted through data mining will enhance data sharing, analyze diversified student relationship management, predict student performances and success of programs etc. Information like what if analysis, predicting modeling, automated alerts, students course tracking, subjects and areas tracking, grade tracking etc. could be effectively and

efficiently enhanced. Using these models one can find which student is likely to be admitted, likely to be enrolled in a particular course, likely to opt electives, likely to default, etc. By addressing these challenges, data mining ensures effective allocation of resources and staff, increases productivity without increasing the cost.

Management institutions are always challenged to stay relevant both in terms of education and research. A quick review of the Financial Times, The Economist, or virtually any magazine or newspaper that covers Management institutions lead anybody to conclude that these institutions are under constant assault by industry, journalists and academics alike to justify their existence, relevancy and effectiveness, given the rapid rate of change in today's world (Sargenti *et al.*, 2006).

The purpose and objectives of the study is:

- To outline a future direction for Management education through emerging applications of data mining and explore the effects of probable changes in recruitments, admissions and course delivery.
- To ensure efficiency in the quality of students, student assessments and evaluations, course evaluations and allocations.

This will be useful to help faculty for managing their classes, understand their students' learning and reflect on their teaching and to support learner reflection and provide proactive feedback to learners. The management institutions collect large amounts of students' data (semantically rich data). Data mining enables management institutions to gain more comprehensive, integrative and reflexive view of the impact of the technologies by obtaining a better understanding around issues of information use and access, ultimately leading to improved knowledge sharing and effective decision making.

BACKGROUND AND RELATED WORK

Data mining is applied to all sectors of the corporates, like retail, banking, telecommunications, marketing, web mining, medicines etc. One can refer to (Adriaans and Dolf, 2005; Fayyad *et al.*, 1996; Han, 2002; Edelstein, 2000; Chang and Lee, 2000; Mobasher *et al.*, 1996; Baylis, 1999; Brossette *et al.*, 1998) for different applications of data mining.

Data mining is widely used in business context but its applications in Management institutions is very much limited. This is the main motivation for present study.

Data mining in education has been attracting recent attention. The concept is relatively recent but number of studies by researchers is going on. The concept is promising but still streamlining is required by relevance of proof and stereotypical analysis. Interesting findings in educational domain include Luan (2002), Petrides and Lisa (2004), Luan (2001), Tsantis and John (2001), Mostow (2004), Merceron and Kalina (2005), SPSS (2005), WEKA (2003), XLMiner (2005) and IlliMine (2006) are available. But these do not address the needs and purpose of management institutions. There are tools (Zaiane, 2001; Merceron and Kalina, 2005) dedicated at finding pedagogically relevant information in student work like TADA-Ed (Tool for Advanced Data Analysis in Education).

One may argue that when statistical analysis is already performed on such educational data, what is the need for data mining. Statistical inference is assumption driven in the sense that a hypothesis is formed and tested against data. Data mining, in contrast is discovery driven. That is, the hypothesis is automatically extracted from the given data. The other reason is that data mining techniques tend to be more robust to both messier real world data and also used by less expert users. The data mining process is very efficient when there are large records for analysis. Delmater and Handcock (2001) expressed that the science underlying predictive modeling

is a mixture of mathematics, computer science and domain expertise. Luan (2001) has produced a high level view of the comparable features among data mining, statistics and data warehouse based on analytical processing.

Tsantis and John (2001) mentions that to take advantage of the results, a system needs to be in place for transforming new knowledge into successful models for teaching and learning to develop and improve student relationship management. Luan (2001) emphasizes on knowledge management for higher education and emphasizes the role of data mining in research, teaching and institutional research.

Techniques in data mining: Feelders *et al.* (2000), Berthold and Hand (1999) and Han and Kamber (2006) identify the data mining process as:

- Definition of the objectives of the analysis.
- Selection and pretreatment of the data.
- Explanatory analysis.
- Specification of the statistical methods.
- Analysis of the data.
- Evaluation and comparison of methods.
- Interpretation of the chosen model.

The techniques and methods in data mining need brief mention to have better understanding.

Associations, mining frequent patterns: These methods identify rules of affinities among the collections. The applications of association rules include market basket analysis, attached mailing in direct marketing, Fraud detection, Department store floor/shelf planning etc. Association can be used to track students activities related to discipline programs, specializations and courses. It was introduced and applied by Agrawal *et al.* (1993) and Agrawal and Srikant (1994). The goal of the association rules is to detect relationships or associations between specific values of nominal attributes in large data sets.

Classification and prediction: The classification and prediction models are two data analysis techniques that are used to describe data classes and predict future data classes. The performance levels of the student can be classified as Good, Medium, or Poor. Using prediction, student's choice of specialization and whether the student will opt one particular course or not can be determined. To predict whether a student will show a certain type of behavior implies an assumption that the student belongs to certain type of student group and will therefore show the certain kind of behavior. The models

of decision trees, neural networks based classifications schemes are very much useful for academic decisions. Decision trees are widely used in prediction and in exploration of datasets like looking at the predictors and values that are chosen for each split of the tree. Regression is often used as it is a statistical method used for numeric prediction. We limit our discussion on algorithms and proof here.

Clustering: It is a method by which similar records are grouped together. Clustering is usually used to mean segmentation. An institution can take the hierarchy of classes that group similar students. Using clustering, students can be grouped based on educational background, age, areas of interest and specialization and so on. The aim of clustering is to create n-groups of students of homogeneous levels with respect to learning.

APPLYING DATA MINING TECHNIQUES TO MANAGEMENT INSTITUTIONS

We provide a conceptual framework for management institutions using data mining techniques (Fig. 1). The framework will address the admission and counseling process with additional help for choice of elective

subjects to students. We are particularly interested in analyzing the data for patterns, such as students' interactions and evaluations, admissions processes and procedures, students' counseling mechanisms, list of courses and subjects offered etc.

Many information technology implementations in management institutions fail not because of technology but because of insufficient attention is paid to issues related to management institution's culture. The approach we have selected is in compliance with the guidelines and steps in the Cross Industry Standard Process for Data mining (CRISP-DM) (Chapman *et al.*, 2000). This process includes understanding the problem and setting the objectives, data preparation, modeling and evaluation as shown in Fig. 2. The framework will try to attempt answers for questions like what types of courses are attracted to what type of students. One can assess the probability of students who would opt for a particular specialization/course. In any students' database, one is not sure about the courses that are taken as a group and which course will be opted by which student.

The proposed framework has four major processes which usually occur in all management institutions namely admissions (planning, evaluation and registration), counseling and allotment of specializations subjects. Each

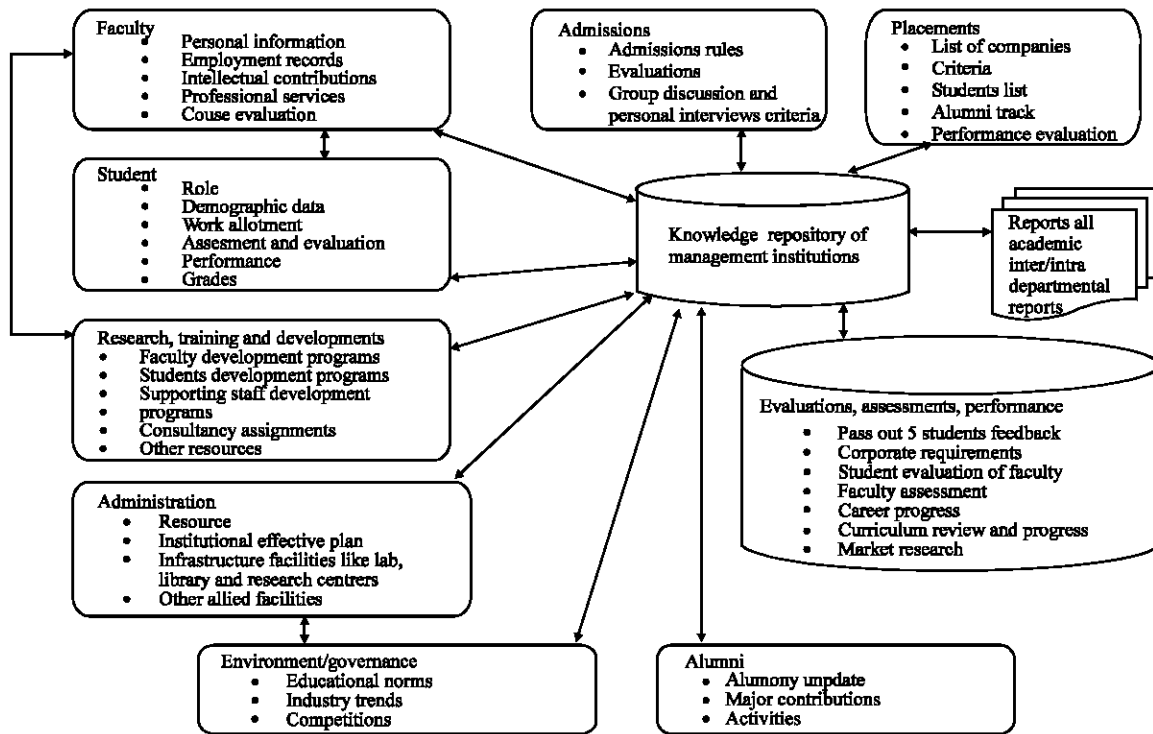


Fig. 1: Conceptual view of resources in management institutions

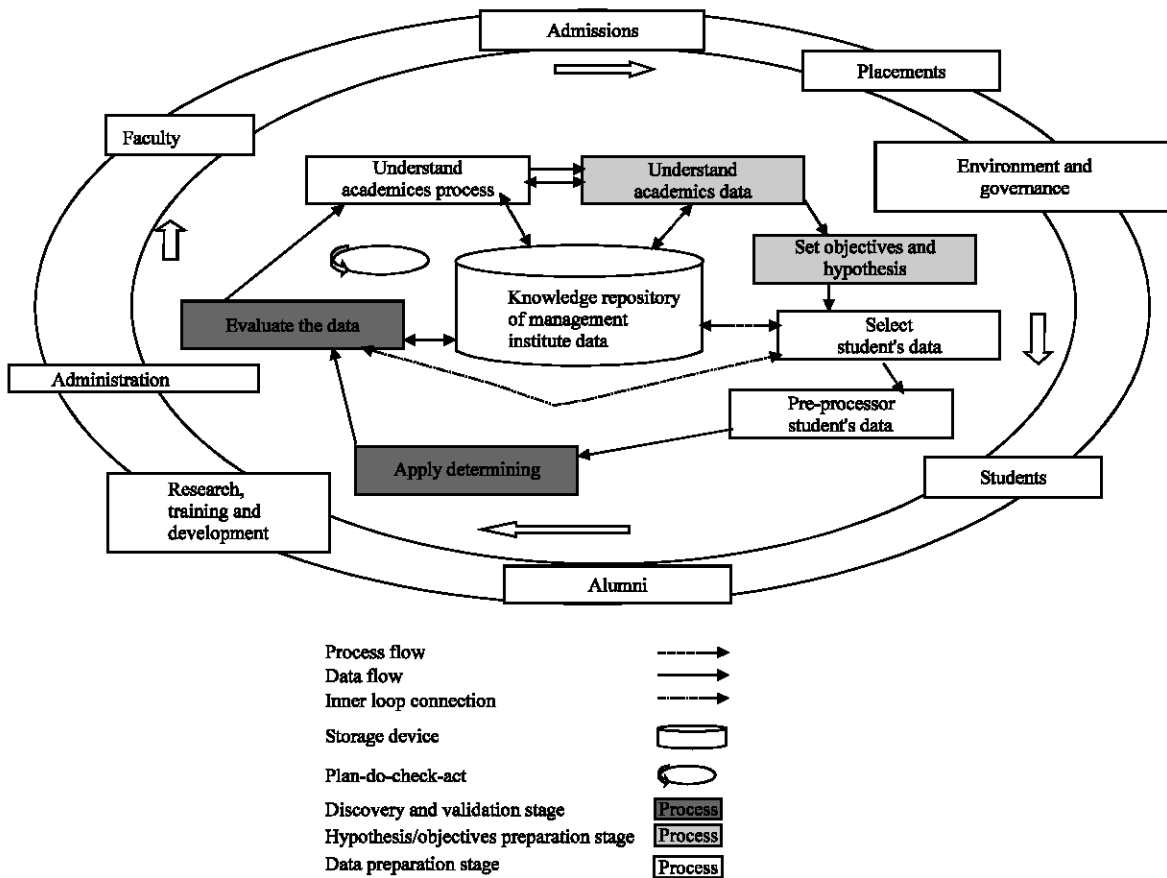


Fig. 2: Data mining process in business school

process can be divided into sub processes. The knowledge that is extracted will be extremely useful for admission process which is very crucial to all institutions.

- The main objective as per the first step in CRISP-DM is improving the admission and counseling process for intake of quality students. Every institute invests on their offerings to nurture students' talents and abilities. Hence our focus is how to best assess the admission and counseling process to improve the quality of students coming to the institute. Since many program specializations and courses are offered to students to select their choice, the proposed framework/model should assist the institute to take decision on which student will opt what type of course.
- The second step is data understanding. This process starts with initial data collection pertaining to students. This phase addresses the data quality problems to discover first insights into the data and to detect interesting subsets to form hypothesis for hidden information. Descriptive Statistics can be

used for data auditing. Balancing the data like enrolled student versus not enrolled students is very much important. The main components here can be students' demographic data, their academic levels and descriptive statistics (relevant without noise and missing values).

- The third step is data preparation. In this the complete dataset is prepared for final data modeling techniques.
- The fourth step is modeling. The conceptual view (Fig. 1) describes the development of knowledge flows in any management institution. This also includes the proactive use of data to assist the administration, faculty and staff in tracking and assessing the performance of all (i.e., students, faculty and staff). The benefits include fast, timely and accurate data, adaptability and reach of any real time management institute data, ability to personalize it, enrolment management, efficient curriculum reviews and program assessments. For building models on admission data, one can use logistic regression, classification and decision trees. For

categorical outcome of admission related data, we can build decision tree or rule sets. For doing the courses/subjects relevance analysis and sensitivity analysis, artificial neural networks can be developed and trained.

- The final step is evaluation and deployment. It is designed to keep track of students' data, credentials and academic results to stream line the process of students' performance. We can apply predictive model in future admissions data sets for recruitment plan, conversion efforts and other administrative decisions based on past historical academic data. One can evaluate whether high quality students have taken admission or not. The prediction model will help Management institutions to identify those students who are more likely to enroll and succeed. The model evaluation includes academic preparedness, demographic information, academic affiliations etc.

Discussion: In Management institutions there are many diverse yet interesting databases available ranging from students, faculty, courses, admin to research and consultancy, infrastructure facilities etc. One may wish to know the information such as which student will opt what specialization during admission and counseling process. Or the student may be interested to know the best course, best program based on prediction of how will they perform in the courses selected. The given framework addresses these types of issues and aids in contributing the increase of quality education delivery. It will also address issues like what type of students are eligible to quality for final placements as per company's requirements.

The admission process: The students get admitted into reputed management institutions based on their performance in Common Admission Test (CAT) conducted by Indian Institute of Managements. The selection is done on the basis of his highest percentile scored in the entrance, Group Discussion (GD) and Personal Interview (PI). The GD includes speaking and discussing in a group of ten to fifteen students on a given topic or case, summarizing and also a five-minute write up of the discussion. The PI majorly focuses on the aptitude, attitude and analytical ability of the student. The over all performance is measured on CAT score, performance of GD and PI, academic background (like Engineering, non-engineering, other special and specific subjects and courses studied), work experience (if-any, ranging from a minimum of six months), extra curricular activities, over all personality and several other factors.

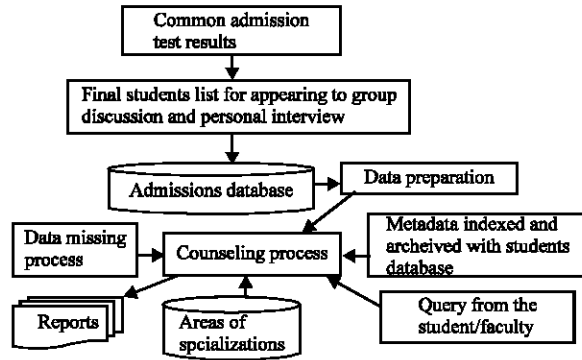


Fig. 3: Counseling process in a business school

Counseling: Among the programs, one can prefer to take International Business, Marketing, Finance, Information Technology, Human Resources or insurance as their specializations. In the admission process, students' demographic and academic data is taken to validate the admission and selection process and effectiveness of the counseling program (Fig. 3). The counseling program starts only after the students are selected in the admission process and get registered. The major consideration to sort out here is there may be too many students taking one particular program specialization which increases the load on both the students and the faculty. This is in fact undesirable. On the other hand, if the right students are missed to a particular program, then the student ends up performing poorly in the examinations and his/her performance in placements affects (the expectations from the faculty and the Business school are very high).

Data mining techniques in admissions and counseling: Decision trees, Bayesian models and other prediction techniques address this admission and counseling process. The suitable data mining method for admissions and counseling that we recommend is based on association rules and prediction. We expect this mechanism, if implemented, will yield promising results. Usually Management institutions prefer cut-off method for selecting students. Instead of assigning each student with a specific program specialization, associative rules based model can be built. That is, assigning a probability estimate to each student to express the likelihood that the student will do poorly in one program specialization. It is however noted that all the examinations usually are evaluated through their aggregated scores. For example, one can predict the likelihood of a student with a background of electronic and communications engineering opting for Human Resources specialization.

What is more important is association rules work on discrete data and hence the students database should be

discretized. It may be possible that only 4 percentage of students apply a particular program specialization and remaining 96 percentages may opt for other specializations. After we get the list of patterns of students and specializations (like which student will opt what specialization) one can use prediction. Through this, one can identify which student is likely to do badly if proper program specialization is not chosen.

Data preprocessing challenges: Students' data quality may be insufficient if data is collected without any specific analysis in mind (Feelders *et al.*, 2000). A careful process with respect to missing or noisy data is required.

To show whether the new specialization program is great and superior to the previous or existing programs for accomplishing the institution's goals, the management institutions need systematic, complete data to evaluate the effects of independent data that are likely to affect new specialization program outcomes like different specializations, dual specializations etc. If the students have not entered certain types of data and if students have ignored thinking that it may not be relevant to them, then it results in missing data and hence interesting hidden patterns that are generated may be incorrect.

Incomplete, noisy and inconsistent data are common when database size is large. There are several reasons for this. The data of students' interest may not be available, or students have not included it simply because it was not considered important at the time of registration. Relevant data may not be recorded due to a misunderstanding. There have been human and computer errors in data entry too. Students not wanting to give details, outdated addresses, too many options for questions, entity identification from multiple sources problem (example: Gates Bill = Bill Gates) are all examples of data discrepancy detections.

Preprocessing allows the user to transform databases to a format that transforms the selected data into a suitable shape to be used by a particular algorithm. Usually, in data mining tools, the transformation performed is a re-casting of the data, or the calculation of a new attribute from existing ones.

EXPECTED BENEFITS

Without an effective way to manage and query the management education data, collected information about students often go underutilized. Parts of a collection can remain untapped for years and the larger it grows, the more difficult its management becomes. Unfortunately, improving this usually comes at a cost-at a time when budget cuts have forced most programs to reduce

spending, any plan for improving management must include how those improvements can increase efficiency and drive down costs in the long run.

The objective of present study is to better understand how and why students chose to study a particular program and what could be done to impact on efficiency. These include:

- Identify barriers in applying to electives.
- Fine tune students' knowledge in their existing database.
- Better manage the performance of students.
- More accurately determine program/specialization quotas (predictive analysis).
- More accurately forecast revenues through these programs for the institution.

The proposed framework improves admission process by identifying the potential students who have the strongest prospects and also forecasts demand for new courses, choice of electives. The benefits include deeper understanding of patterns previously unseen using current available reporting capabilities. Further prediction allows the institution an opportunity to act before students withdraw a specialization or an elective course. The institutions can know the resource allocation with confidence gained from knowledge how many students will take a particular course.

Data mining allows the institute to keep track of all the transactions that can be queried at any given time the student transactions that need immediate attention can be visualized by charts and graphs. Patterns include targeting a group of students who may need special attention, monitoring consistent high grades, monitoring attendance, discipline referrals or any combination of student indicators. This way, it saves time, reduces staff burn out and improved delivery of educational services. It enhances the data sharing by efficiently using the institute's resources. Table 1 shows certain patterns that can be generated using data mining.

Focusing on student retention in some specializations/programs, can be done by predicting the accuracy of data mining's decision trees and artificial neural networks with that of logistic regression. Questions related to the registration behaviors of the admitted students in management institutions can be explored like this: Do admitted students enroll for each program/specialization randomly? Are certain admitted students more likely to enroll in some program/specialization than others? Data-mining modeling processes can be adopted here and evaluated in comparison to the traditional logistic regression approach.

Table 1: Patterns generated through data mining process queries

Major data mining techniques	Patterns generated through Data mining Process Queries
Clustering	Students having similar characteristics. Grouping top performers, average performers, under performers. Similar Segments related to marks, grades, disciplines, performance and training.
Classification and prediction	Predicting student learning outcome in a program/course. Predicting the likelihood of success in a program/course. Predicting the percentage of accuracy in students' performance. Analyze forecast and model information to quantify faculty and student assets. Classifying the Admissions in students recruiting process. Prediction of what type of students most likely to take a particular program/course, most likely to shift, most likely to take particular assignments. Predicting student's behavior, attitude. Predicting demand projections of students by seasonal variations. Predicting the Performance progress throughout the program year. Identifying the best profile for different students. Prediction to find what factors will attract prospective students. Classify trends of movements through the organization for successful/unsuccessful students in placements.
Association	Association of training/projects undertaken with various types of students and performance scores, individually and in teams. Association of students' work profiles to the most appropriate programs/jobs Association of students' team building and leader ships approaches. Association of students' attitude with performance. Association and analysis of students' movements Application tracking mechanism in admission process.
Data mining using other inter-disciplinary methods	Standardizing training methods, task performance monitoring in students' career management, monitor work flow route tasks. Provides summary information: various multidimensional summary reports. Statistical summary information (data central tendency and variation). Use historical data to build models of fraudulent behavior and use data mining to help identify similar instances.

LIMITATIONS

Data mining applied to large management institutions can be used to monitor the budget to actual performance. The institutes can tailor and monitor the fund raising strategies. It improves the admission process by identifying the potential students who have the strongest prospects and also forecasts demand for new course, choice of electives. It may be slightly premature to analyze the data of management institutions in India as many of them have been established recently. Institutions who spend more than 15 years imparting management education with technology infrastructure, it is time for them to turn their attention in improving the quality of education. The students are increasing their demand consciously for information and analysis of each management institution.

FUTURE WORK

We intent to further explore in future a series of factors relating to institute size, type and technology to see if there is any relationship between these factors and use of data mining methods. The future work may include the creation of a data warehouse which includes courses, programs and curriculum folders, with predefined relationships established which will allow faculty and staff to track student and faculty performance over time in a real time manner with a richer data set. No longer one has to search several databases for student information. We

also plan to make it as easy as possible for teachers not too familiar with new technologies. Finally, we would like to develop this tool in a more intelligent manner, to improve the quality of teaching.

CONCLUSION

It is clear that recognition of high quality educational research by professors and staff is dealt for students admissions every year. Since the number of seats have significantly grown over the period of years, it becomes prudent to look at how teaching and learning have changed. The institutes have been making substantial investments for their computing infrastructure to meet their goals. All institutes are using the information about their students to gain insights into bigger issues like students' performance, placements, students' admissions, students' successes and performances. The regulatory bodies, accreditation bodies are seeking more information to measure and evaluate the effectiveness of the institutes often termed as ratings. In this paper, we have presented the conceptual framework in the adoption of data mining in Management institutions. This will be useful to help faculty for managing their classes and courses, understand their students' learning and provide proactive feedback to learners. The impetus of prevalent advanced applications like Predicting student performances and success of programs, automated alerts and what if analysis using data mining will be extremely helpful in academic learning. Data mining intersects technology,

information, management practices through efficient models, analyzes the diversified student relationship management, assesses the alumni affairs etc.

REFERENCES

- Adriaans, P. and Z. Dolf, 2005. Data mining. Pearson Education, pp: 69-71.
- Agrawal, R., T. Imielinski and A. Swami, 1993. Mining association rules between sets of items in large databases. Proceedings of SIGMOD-93, Retrieved April 1, 2006, from <http://rakesh.agrawal-family.com/papers/sigmod93assoc.pdf>.
- Agrawal, R. and R. Srikant, 1994. Fast algorithms for mining association rules. Proceeding of VLDB-94. Retrieved on January 31, 2000, from <http://www.sigmod.org/vldb/conf/1994/P487.pdf>.
- Baylis, P., 1999. Better health care with data mining, SPSS White Paper, UK. Retrieved March 18, 2005, from <http://www.sinter.com.tw/spss/documents/Better%20health%20care%20with%20data%20mining.pdf>.
- Berthold, M. and D.J. Hand, 1999. Intelligent Data Analysis: An Introduction. Springer, pp: 3-10.
- Brossette, S.E., A.P. Sprague, J.M. Hardin, K.B. Waites, W.T. Jones and S.A. Moser, 1998. Association rules and data mining in hospital infection control and public health surveillance. *J. Am. Med. Inform. Assoc. (JAMIA)*, 5: 373-381.
- Chang, W.H. and Y.H. Lee, 2000. Telecommunications data mining for target marketing. *J. Comput.*, 12: 60-74.
- Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth, 2000. CRISP-DM 1.0: Step-by-step data mining guide. Retrieved March 14, 2006, from www.crisp-dm.org/CRISPWP-0800.pdf.
- Delmater, R. and M. Handcock, 2001. Data mining explained: A Manager's Guide to Customer-Centric Business Intelligence, Digital Press. Boston, MA.
- Edelstein, H., 2000. Building profitable customer relationships with data mining. SPSS white paper-executive briefing. Retrieved March 14, 2006, from http://www.spss.fi/PDF/Building_profitable_cust_relations_DM.pdf.
- Fayyad, U.M., G. Piatsky-Shapiro, P. Smyth and R. Uthurusamy, 1996. *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: AAAI Press.
- Feelders, A., H. Daniels and M. Holsheimer, 2000. Methodological and practical aspects of data mining. *Inform. Manage.*, pp: 271-281.
- Han, J., 2002. How can data mining help bio-data analysis. Workshop on data mining in bioinformatics BIOKDD02. Retrieved September 14, 2005 from <http://www.cs.rpi.edu/~zaki/BIOKDD02/01-han.pdf>.
- Han, J. and M. Kamber, 2006. *Data mining concepts and techniques*. Morgan Kaufmann, pp: 4-27.
- IlliMine, 2006. IlliMine software, version 1.1.1., Retrieved September 13, 2006, from <http://illimine.cs.uiuc.edu/>.
- Luan, J., 2001. *Data Mining Applications in Higher Education: A Chapter in the Upcoming New Directions for Institutional Research*, Josse-Bass. San Francisco.
- Luan, J., 2002. Data mining and knowledge management in higher education-potential application. Proceedings of Association of Institutional Research (AIR) Forum, Toronto, Canada.
- Merceron, A. and Y. Kalina, 2005. *Educational Data Mining: A Case Study*. Artificial Intelligence in Education: Supporting Learning Through Intelligent and Socially Informed Technology. Proceedings in Frontiers in Artificial Intelligence and Applications *et al.* IOS Press.
- Mobasher, B., N. Jain, E. Han and J. Srivastava, 1996. Web mining: Pattern discovery from world wide web transaction. Technical Report TR96-050, Department of Computer Science, University of Minnesota.
- Mostow, J., 2004. Some useful design tactics for mining ITS data. Proceedings of ITS2004 Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes, Maceio, Brazil.
- Petrides, A. and Lisa, 2004. Knowledge management, information systems and organizations. ECAR: Centre for Applied Research. Res. Bull., No. 20.
- Sargenti, P., L. William and K. Mounir, 2006. Diffusion of knowledge in and through higher education organizations. *Issues Inform. Syst.*, 3: 312-316.
- SPSS, 2005. Clementine. Retrieved March 13, 2005, from www.spss.com/clementine.
- Tsantis, L. and C. John, 2001. Enhancing learning environments through solution-based knowledge discovery tools: Forecasting for self-perpetuating systemic reform. *J. Special Edu. Technol. E- J.*, 16.
- WEKA, 2003. Retrieved February 14, 2006 from www.cs.waikato.ac.nz/ml/weka.
- XLMiner, 2005. XLMiner for Windows. Retrieved February 14, 2005, from <http://www.resample.com/xlminer>
- Zaiane, O.R., 2001. Web usage mining for a better web-based learning environment. Proceedings of Conference on Advanced Technology for Education (CATE), Banff, Alberta, pp: 60-64.