'National Conference on Recent Advances of Computational Intelligence
Techniques in Science, Engineering and Technology'
International Journal of Scientific Research in Computer Science,
Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Data Mining In Higher Education- Persistence Clustering and Prediction

Dr. R. Senkamalavalli[1], Shammi.L[2]
[1]Associate Professor, [2]Assistant Professor
Department of CSE, East Point College of Engineering and Technology, Bangalore, Karnataka, India

ABSTRACT

Data mining is slowly but surely making its way into the educational field after dominating the business fields. Higher education will find larger and wider applications for data mining than its counterpart in the business sector, because higher education institutions carry
their duties that are data mining
Intensive: scientific research that relates to the creation of knowledge, teaching those concerns with the transmission of knowledge, and institutional research that pertains to the use of knowledge for decision making. All the above tasks are well within the boundaries of knowledge management, which drives the need for better and faster decision-making tools and methods.

Data mining enable 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.

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.

Keywords: Decision Making, Higher Education, predictive Model.

I. INTRODUCTION

One of the biggest challenges that is faced by higher education today is predicting the paths of students and alumni institutions would like to know, for example, which students will enroll in particular course programs, and which students will need assistance in order to get placements etc.? In addition to this challenge, traditional issues such as management, financial assistance and time- to-degree continue to motivate higher education institutions to search for better solutions.

One way to effectively address these students and alumni challenges is through the analysis and presentation of data, or data mining.

II. RELATED WORK

In 2007, Vranice et al., in [2] explored data mining algorithms on a Croatian university students' data. The focus was whether future students of this course will succeed or fail. They have tested several algorithms and their result was similar, however, the authors indicated that their sample was small and perhaps future research would include more detailed student data.

Parack et al., in [3] presented a paper focusing on predicting academic trends and student patterns behavior. This has eased the process in grouping similar student profiles and identifying their learning patterns.

Very interesting paper by Nasiri and Minaei in [4] in higher education focused on two issues: GPA and academic dismissal in a Learning Management System (LMS). Both algorithms used for the data mining process indicated a weakness if there is a slight variation to the data, this will lead to different results, they solved it by adding association rules.

Shi et al., presented a paper [5] focusing on managing the university curricula based on data association mining technology. They reached the conclusion that if a student was successful in a certain course, then he will be Successful in similar courses as well. An example can be mathematics and physics.

Again in India, Bunkar et al., presented a paper where they applied data mining techniques to predict the performance improvement of graduate students using classification [6]. Several techniques were discussed and the authors were able to isolate students that are most likely to fail and provide proper counseling and guidance.

In Romania, Bresfelean et al., exploited the university academic failure issue in [7]. Their aim was to define an academic failure profile for students to be able to predict students' exam failure and success based on data mining techniques. They aimed to improve students learning methods and detect their weakness, and assist in managerial educational decision.

Offering high quality education means being able to predict student enrollment in courses, identify beneficial teaching methods, forecast student performance in end exams and identify drop out rates, and help those students during the semester. Their method described in [8] based on classification helped in the proper dividing of students, and paying special attention to students most likely to fail, and help in increasing the success in the success and failure ration.

In China, Wu had a different intake on higher education by using clustering to identify student course selection based on teachers [9]. Their goal was based on guiding the students and giving them most appropriate advice to succeed. The author identifies different categories of student-teacher selection, out of the three clusters, one was successful in selecting the teacher based on several important factors which has increased their class interactivity, discipline, behavior and led to their success. The other two groups, should receive proper guidance to achieve what the first group has achieved.

Japan had another take on the topic by focusing on the university curricula and built an EDM to reach an optimal learning success in terms of best possible Grade Point Average (GPA) [10]. They enforced their system by including an individual learner profile that includes pre- university educational data and grows as the student progresses through the university. This helps in grouping similar profiles and inferring success patterns that can't be identified through conventional student analyzing methods.

A recent study in 2012 in [11] focused on predicting the drop-out rate of students from universities, colleges, and institutions in developing countries. Mustafa et al., used classification and regression to classify successful and unsuccessful students based on gender, financial condition, ethnicity, work-status, disability, and study environment. The search was based on background information. And they identified the most important classification factors were: financial support, age group, and gender.

In 2005, a paper presented by Delavari et al., developed an educational data mining model to be tested in a Malaysian University [12]. They had three targets, the first was understanding the course enrollment pattern in a course, and identifying which students are successful in passing the course, and who will fail. This leads to their next target, faculty, who take proper action in guiding these students, and choose direct or indirect methods to provide the necessary class skills, and financial aid. The final target of the system, are the decision makers, the system enhances the education quality and provide quality management, improve policy making, and setting new strategies and goals.

More on the topic of student failure in college courses can be found in [13]. The author used association rules mining algorithm to find the factors that lead to student course failures.

In Spain, a research presented by Tovar and Soto to improve their predicting model [14]. Instructors can locate students having problems with the course and help them. It also identifies student who have the capabilities to pass the course but fail. Their research is based more on statistics than data mining techniques.

A paper by Knauf et al., [15] presented their storyboard model which students at the Tokyo Denki University were using for progressing through the curriculum. Then they used data mining techniques to build an individual student profile to record their personal properties, talents, weaknesses, and preferences. This model presents students with suggested courses where they will be successful.

Ningning presented a paper focusing on data warehousing and data mining [16]. His model identifies students most likely to drop out a course. This can aid business managers in pinpointing students in need of help and guidance.

III. KNOWLEDGE MANAGEMENT DRIVING DATA MINING

Data mining should be performed on very large or raw datasets using either supervised or unsupervised data mining algorithms.

Several authors have written about the factors behind the down of data mining. For instance, Sir.Therling identified three reasons viz;

- The ease of data collection and storage
- The computing power of modern processors and
- The need for fast and real time data mining.

Yet one important reason absent from these is the growing interest in knowledge management. Knowledge, a focal point of ontology or epistemology, is the product of moving from data to information and finally to knowledge.

IV. DATA MINING MODELS

The following model, tiered knowledge management model (TKMM) developed by Jing Luan illustrates the dichotomous nature of modern knowledge management framework for higher education research professionals.

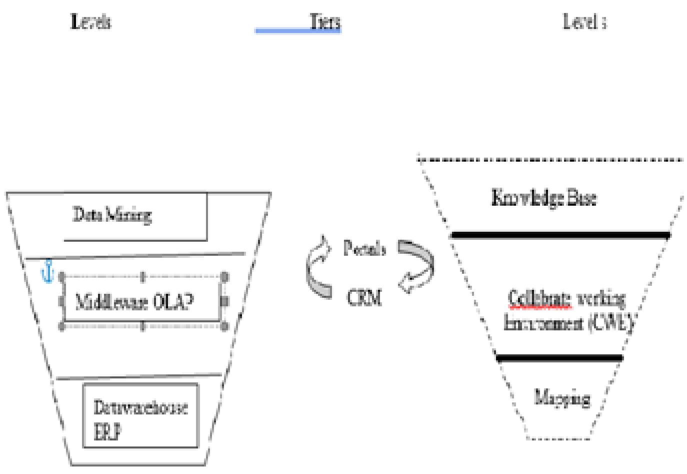

Figure 4.1Tiered Knowledge Management Model (TKMM)

Models house the steps modules, and resources of the data mining process. Some data mining models include the entire process for a particular purpose, be it to cluster or predict.

Data mining is a powerful tool for academic intervention. The components of knowledge management are explicit (documented, measurable) and tacit (subjective, qualitative).

Documented, measurable explicit knowledge is most familiar and available to us, as it exists mostly in databases and other similar medium. While tacit knowledge, an entity of feelings, personalities and aptitudes is crucially important but it is hard to quantify and we will leave that for further study. All these components in customer relationship management (CRM) operational, analytical and collaborative are key users of data mining.

On the explicit side, data mining reflects the highest level of knowledge attainment that requires skills in data domain (tier one). Data querying and presentation (tier two). And artificial intelligence /machine learning (tier three). Data mining occupies the top tier and is dependent on the lower tiers. The following chart is a topology of the explicit knowledge of TKMM that illustrative in detail the relationships among three tiers and the software programs for each:

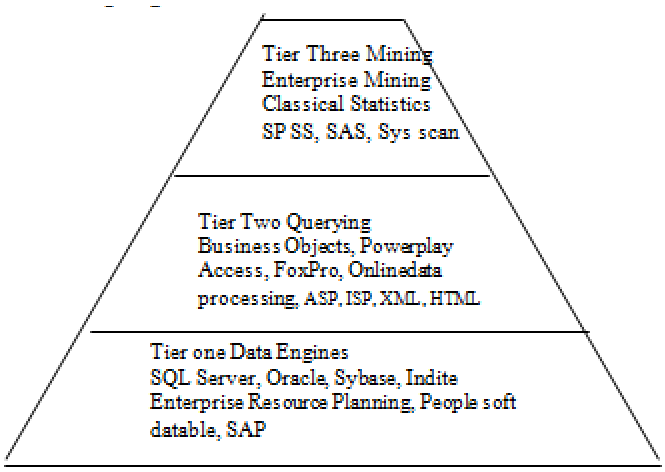

Figure 4.2: Topography of Tiered Knowledge Management Model (TKMM) for explicit Knowledge

4.1.1 SUPERVISED AND UNSUPERVISED MODELING

Classification and estimation use either unsupervised or supervised modeling techniques. Unsupervised data mining is used for situations in which particular groupings or patterns are unknown. In student course databases, for example, little is known about which courses are usually taken as a group, or which courses types are associated with which student types. Unsupervised data mining is often used first to study patterns and search for previously hidden patterns, in order to understand, classify, typify and code the objects of study before applying theories. Supervised data mining, however, is used with records that have a known outcome. A graduation database, for example, contains records of students who completed their studies, as well as of those who dropped out. Supervised data mining is used to study the academic behavior of both groups, with the intention of linking behavior patterns to academic histories and other recorded information.

V. DATA MINING IN HIGHER EDUCATION

Data mining is already fundamental to the private sector. Many of the data mining techniques used in the corporate world, however, are transferable to higher education.

What are the transferable techniques in data mining that are readily applicable in higher education? Infact, there are many algorithms are similar in concept to stored producers in object related programming in that they are universally applicable.

The following case study illustrates a key application of data mining in higher education.

Case Study: **Creating meaningful learning outcome typologies**

"What do institutions know about their students?" If the answer is a recital of enrolment percentages of the basic counts, institutions do not know their students as well as they could. This case study demonstrates how suburban community colleges can establish learning outcome typologies for students using unsupervised data mining.

SOLUTION

To establish appropriate typologies, we can use two step and k-means clustering algorithms we first will apply the algorithms to the general groupings that identifies students as "transfer oriented", "basic skill upgraded", " on campus directed", "vocational education directed" postal course oriented " etc, Note that the boundaries among clusters were under and dispersed.

Now we need to have a replacement method that takes care of cases, which do not appear to or belong to any group. Defining educational outcomes is easier said that done. Dropping out is also an outcome by itself. Further we need to determine the length of study which required decisions on how to deal with 'stop outs' students who left colleges and later returned. All of these situations test the data miners domain knowledge.

The two-step algorithm will produce following clusters: "Transfers", "Vocational students", "On campus directed", "Basic skills students", "students with mixed outcomes", and "Drop outs". K-means will validate this cluster. Some transfer students can complete their studies quickly some vocational can take it longer falls within this.

VI. RESULTS

If data mining can quickly identify potential donors by a ratio of two to four (correctly predicting two out of four who will donate) then the university can achieve results

By mailing only to the indicated 40% of the alumni donor population thus saving considerable time and money.

VII. CONCLUSION

With the ability to uncover hidden patterns in large databases, colleges and universities can build models that predict with a high degree of accuracy the behavior or several clusters. By acting on these predictive models educational institutions can effectively address issues ranging from transfers and retention to marketing and alumni relations. Data mining is a new type of exploratory and predictive data analysis that has tremendous applications in higher education institutional research alone. An effective data mining cycle was presented to clarify the process. We will take this research further by incorporating RULES-3 developed by Pham and Aksoy in [17] and later improved by Mathkour in [18], in the Edu Gate System and develop a tool for management to make their insightful decisions. Moreover, we will also focus on data selection, and preparation phases when the system is tested for accuracy, validation, and verification.

VIII. REFERENCES

[1]. D. A. Alhammadi and M. S. Aksoy, "Data Mining in Education – An Experimental Study," International Journal of Computer Applications, vol. 62, no. 15, pp. 31-34, 2013.

[2]. M. Vranic, D. Pintar and Z. Skocir, "The use of data mining in education environment," IEEE Telecommunications ConTel 2007. 9th International Conference, pp. 243-250, 2007.

[3]. S. Parack, Z. Zahid and F. Merchant, "Applications of Data Mining in Educational Databases for Predicting Academic Trends and Patterns," IEEE International Conference on Technology Enhanced Education (ICTEE), pp. 1-4, 2012.

[4]. M. Nasiri and B. Minaei, "Predicting GPA and Academic Dismissal in LMS sing Educational Data Mining: A Case Mining, " IEEE Third International Conference on E- Learning and E-Teaching (ICELET), pp. 53-58, 2012.

[5]. F. Shi, Q. Miao and D. Mei, "The Application of Data Association Mining Technology in University Curriculum Management," IEEE Symposium on Robotics and Applications (ISRA), pp. 521-524, 2012.

[6]. K. Bunkar, U. K. Singh, B. Pandya and R. Bunkar, "Data Mining: Prediction for Performance Improvement of Graduate Students using Classification," IEEE Wireless and Optical Communications Networks (WOCN), pp. 1-5, 2012.

[7]. V. P. Bresfelean, M. Bresfelean, N. Ghisoiu and C.A. Comes, "Determining Students' Academic Failure Profile Founded on Data Mining Methods," IEEE 30th International Conference on Information Technology Interfaces, pp. 317-322, 2008.

[8]. S. K. Yadav, B.K. Bharadwaj and S. Pal, "Data Mining Applications: A comparative study for Predicting Student's Performance", International Journal of Innovative Technology and Creative Engineering (IJITCE), vol. 1, no. 12, pp. 13-19, 2011.

[9]. F. Wu, "Apply Data Mining to Students' Choosing Teachers Under Complete Credit Hour," IEEE Education Technology and Computer Science (ETCS), vol. 1, pp. 606-609, 2010.

[10]. M. N. Mustafa, L. Chowdhury and M. Kamal, "Students Dropout Prediction for Intelligent System from Tertiary Level in Developing Country," International Conference Informatics, Electronics & Vision (ICIEV), pp. 113-118, 2012.

[11]. N. Delavari, M. R. Beikzadeh and S. Phon-Amnuaisuk, "Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System," IEEE International Conference on Information Technology Based Higher Education and Training (ITHET), pp. F4B/1-F4B/6, 2005.

[12]. Z. Zhang, "Study and Analysis of Data Mining Technology in College Courses Students Failed," International Conference on Intelligent Computing and Integrated Systems (ICISS), pp. 800-802, 2010.

[13]. E. Tovar and O. Soto, "The Use of Competences Assessment to Predict the Performance of First Year Students," IEEE Frontiers in Education Conference (FIE), pp. F3J-1, 2010.

[14]. R. Knauf, Y. Sakurai, K. Takada and S. Tsuruta, "Personalizing Learning Processes by Data Mining," IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 488-492, 2010.

[15]. G. Ningning, "Proposing Data Warehouse and Data Mining in Teaching Management Research," International Forum on Information Technology and Applications (IFITA), vol. 1, pp. 436-439, 2010.

[16]. D. T. Pham and M. S.Aksoy, "RULES: A Simple Rule Extraction System," Expert Systems with Applications, vol. 8, no. 1, pp. 59-65, 1995.

[17]. H. Mathkour, "RULES3-EXT: Improvements of RULES3 Induction Algorithm," Mathematical and Computational Applications, vol. 15, no. 3, pp. 318-324, 2010.