ISSN 2303-4521
PERIODICALS OF ENGINEERING AND NATURAL SCIENCES
Vol. 1 No. 2 (2013)
Available online at: http://pen.ius.edu.ba

Data Mining in Higher Education

Dina A.Aziz AlHammadi
Department of Information Sciences
College of Computer and Information Sciences
King Saudi University P.O.Box 2454, Riyadh 11451
Saudi Arabia
dina.alhammadi@gmail.com

Mehmet Sabih Aksoy
Department of Information Sciences
College of Computer and Information Sciences
King Saudi University P.O.Box 2454, Riyadh 11451
Saudi Arabia
msaksoy@ksu.edu.sa

Abstract
Data mining is slowly but surely making its way into the educational field after dominating the business fields. This paper focuses on the research completed in the area of data mining in the higher education sector: colleges and universities. We will look at the different implementation of data mining and to what extent was it utilized and benefited from.

Keywords: data mining, higher education, inductive learning, rules-3

1. Introduction

Data mining is now the tool to advance every education establishment in every country. And now many nations are using data mining to upgrade their education, and thus producing highly educated citizens who can elevate the other sectors, and upheave the economy and value of the nations.

Data mining is now considered a great asset to the educational sector in highlighting and scrutinizing many elements absent from educational managers. An experimental study was presented in [1]. Our scope is higher education, and the managers can be any professor, instructor, academic counselor, teacher, head, teaching assistant, etc. Providing certain data to these managers can help in early identification of students who are falling behind their peers, guiding students through the best college curricula path, deciding majors, identifying pass and fail rations, locating education related issues and the ability to sole these problems early on in the learning process.

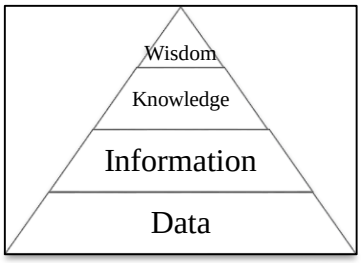

Figure 1. DIKW Pyramid

In Figure 1, Data is unprocessed and raw facts that hold no meaning, as we move to the next layer, the data is now information, this means that data has undergone some processing to produce meaningful information, the next layer is mining the information for knowledge, and moving to the top layer means we have managed to acquire wisdom. The main text of this study consists of three sections. The second section is the work covered in the area of educational data mining. The third section is the conclusion.

2. 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 algorithm 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

DOI: 10.21533/pen.v1i2.17

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.

3. Case Study

An important question to ask higher education institutes is "How much do you know about your students?" And if the answer is the number of enrolled students and percentages of those who passed and failed, then this the wring answer. The right answer would be acknowledging the student's choices in curriculum and college life. This includes finance, achievements, success, and tailored guidance by their academic counselors.

Higher education institutes are in great need of embedding data mining tools as they are essential for progressing academically.

We will explore three different cases where data mining is essential and we will show its benefits.

3.1. Case 1: Students' Level

Students are the basic building block of any educational institute. In any higher education institute students are somewhat dependent on themselves to chooses their own

courses, professors, make important decisions such as finances, dropping from the university, changing majors, and transfers.

Sometimes, students choose courses based on their peers, past experiences with certain professors, and other related or directly unrelated issues. This has lead to the fact that students make wrong or missed choices.

The suggested solution would be incorporating data mining in the educational gate which provides many services to students at King Saud University such as but not limited to: course registration, dropping course, appealing for postponing a semester, and finance.

If based on students' grades, past courses passing or failing, personal background, we may find a pattern where other students were successful, and thus recommend courses and decisions that are likely to help the students achieve much more than he already gaining from his university education. Students will be able to make decisions based on reliable information.

3.2. Academic Guidance Level

Going up the ladder, academic counselors need certain background information to help aid their students. Such as those who are failing courses, or barely passing, constantly transferring between majors, amongst many other issues. If counselors have social, financial and academic background knowledge about the students then they can make a well-informed guiding decision supported by similar past students experiences. Data mining, if incorporated, in the educational gate system of King Saud University will benefit academic counselor by providing essential factors affecting student's success or failure, and will be able to suggest best academic and financial, social paths based on similar patterns by previous students.

3.3. Decision Makers Level

The top of the pyramid is as important in any educational institute as the previous building blocks if not more essential for the success of any higher education institute. Deans, rectors, vice rectors, deanships, college deans, and administrative departments, are all required to make important decisions at any given time. Therefore, information must be delivered to them accurately and in a timely manner. Until recently, decisions are made based on past experiences, however, the institutions can't wait for every decision maker to gain experiences of 10 years or more before having the ability to make the right judgment or gain the ability to assess the institutes capabilities and opportunities. This leads us to incorporating data mining as tools providing not only simple enrollment, transferring or dropping students' data as well as courses and professors. The tool should be able to assess and present the leaders with related data, important factors leading to success or failure regarding students, courses, and professors. Leaders will have an eagle-eye view of the institute, highlighting problem areas, and significant success areas as well. This will be based on past data of all important building blocks and this will help suffice for the experience young leaders are lacking, as identified by the management.

4. Discussion

As portrayed in the three cases, we are not looking at a pyramid, but a cycle where elements extracted from each phase affects the out come of the next phase.

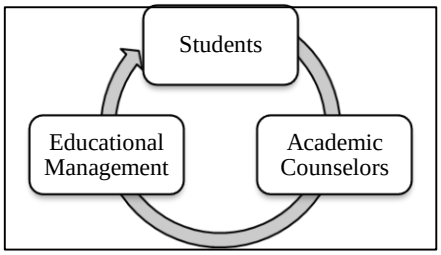

Figure 2: Proposed Data Mining Effect Cycle in Education

Figure 2 clearly explains it. This can be described simply with an example. If we have a student named Ahmad, he is a first year students at the College of Computer and Information Sciences at King Saud University. His first step is registering courses like his peers. The Edu Gate should display to him the required core and elective courses, however, based on his Prepatory Year Program (PYP) courses and other factors where selected patterns are found between him and previous first years students, courses where his success rate is estimated to be high will be suggested to him. If Ahmad registers these suggested courses, and as the system predicted, passes all courses except one. Then we have fed into the system new facts about Ahmed and perhaps a new pattern was introduced. The next year, Ahmad may or may not visit his academic counselor, still, all counselors will have certain facts about troubled students who didn't fair well the previous semester presented in the Edu Gate, alerting the counselor to take notice and act upon the data presented. He may also feed into the system new facts about Ahmed, his courses, or his professors. And because Ahmed failed one course, the system will alarm the counselor to take action based on his PYP data, and passed courses, amongst other important factors highlighted by data mining. And any decision made by the counselor will affect future academic choices presented and decided for Ahmed.

The third phase is laying the most important information to management. Looking more at the big picture rather than unrelated details. Data mining tool in this case will represent problem areas or areas requiring attention, as well as areas that are improving since last assessment. The college dean will notice that there is a course with a high percentage of failure regardless of which students failed. The system will present similar patterns shared by those students that may be a factor. For example; if they are all

taking a course by a single professor, or they have similar weakness area such as: low math grades in PYP, financial issues, or social related issues. Management will be able to focus on those problematic areas and solve the issues instead of making unknowledgeable decisions or neglect those problematical areas. This way the education management will focus their time and effort to improve the institute from all perspectives: students, courses, and faculty. The management will make several changes to all levels, thus feeding back to the students. This is why it should be considered a cycle instead of a pyramid.

5. Conclusions

The aim of this paper is to discuss the need for data mining in all levels of any educational institute. 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.

Acknowledgements

The authors would like to thank Research Center in the College of Computer and Information Sciences, King Saud University for its support to complete this study.

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] Y. Sakurai, K. Takada, S. Tsuruta and R. Knauf, "A Case Study on Using Data Mining for University Curricula," *IEEE 12th International Conference on Advanced Learning Technologies (ICALT)*, pp. 3-4, 2012.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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