Application of Higher Education System for Predicting Student Using Data mining Techniques

P.Veeramuthu   Dr.R.Periasamy
Research scholar   Associate Professor
                                                PG and Research Department of Computer Science
                                                Nehru Memorial College

Abstract— *The aim of research paper is to improve the current trends in the higher education systems to understand from the outside which factors might create loyal students. The necessity of having loyal students motivates higher education systems to know them well, one way to do this is by using valid management and processing of the students database. Data mining methods represent a valid approach for the extraction of precious information from existing students to manage relations with future students. This may indicate at an early stage which type of students will potentially be enrolled and what areas to concentrate upon in higher education systems for support. For this purpose the data mining framework is used for mining related to academic data from enrolled students. The rule generation process is based on the classification method. The generated rules are studied and evaluated using different evaluation methods and the main attributes that may affect the student's loyalty have been highlighted. Software that facilitates the use of the generated rules is built which allows the higher education systems to predict the student's loyalty (numbers of enrolled students) so that they can manage and prepare necessary resources for the new enrolled students.*

Keywords— *Higher education, Data mining, Knowledge discovery, Classification, Association rules, Prediction, utlier analysis.*

I. INTRODUCTION

Nowadays there is an evolution of educational systems and there is a great importance of the educational field. Modern educational organizations start developing and enhancing the educational system increasing their capability to help the decision makers obtain the right knowledge, and to make the best decisions by using the new techniques such as data mining methods [1]. Subsequently, a suitable knowledge needs to be extracted from the existing data. Data mining is the process of extracting useful knowledge and information including: patterns, associations, changes, anomalies and significant structures from a great deal of data stored in databases, data warehouses, or other information repositories. The data mining expediency is delivered through a series of functionalities such as outlier analysis, evolution analysis, association analysis, classification, clustering and prediction. Data mining is an integral part of Knowledge Discovery in Database (KDD) [2][3]. Student enrollment process in any higher education system is of great concern of the higher education managements. Several factors may affect the student enrollment process in a particular Institute. One of the biggest challenges that higher education faces today is predicting the paths of loyal students (enrolled students). Institutions would like to know, for example, which students and how many will enroll in particular institute. This research paper is an attempt to use the data mining processes, particularly predictive classification to enhance the quality of the higher educational system to increase numbers of loyal students (enrollment students) to evaluate student data to study the main attributes that may affect the student enrollment factors to plan for institutes resources (Instructors, classes, labs, etc.) From knowing how many students will be enrolled and to make a big effort to concentrate on all factors that play main role in motivating the new student to enrolls in a particular institute.

II. RELATED WORK

An easy Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes.
 [1] Have case study of using educational data mining in Module course management system. They have described how different data mining techniques can be used in order to improve the course and the students' learning. All these techniques can be applied separately in a same system or together in a hybrid system.
[2] Have a survey on educational data mining between1995 and 2005. They have compared the Traditional Classroom teaching with the Web based Educational System. Also they have discussed the use of Web Mining techniques in Education systems.
[3] Have a described the use of k-means clustering algorithm to predict student's learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students.
[4] Discuss how data mining can help to improve an education system by enabling better understanding of the students. The extra information can help the teachers to manage their classes better and to provide proactive feedback to the students

[6] Have described the use of data mining techniques to predict the strongly related subject in course curricula. This information can further be used to improve the syllabi of any course in any educational institute.

[7] Describes how data mining techniques can be used to determine the student learning result evaluation system is an essential tool and approach for monitoring and controlling the learning quality. From the perspective of data analysis, this Paper conducts a research on student learning result based on data mining.

III. RESEARCH OBJECT

The object of the present study is to identify the potential areas in which data mining techniques can be applied in the field of Higher education and to identify which data mining technique is suited for what kind of application.

IV. DATA MINING DEFINITION AND TECHNIQUES

Simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data. [5] Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. The sequences of steps identified in extracting knowledge from data are: shown in Figure 1.
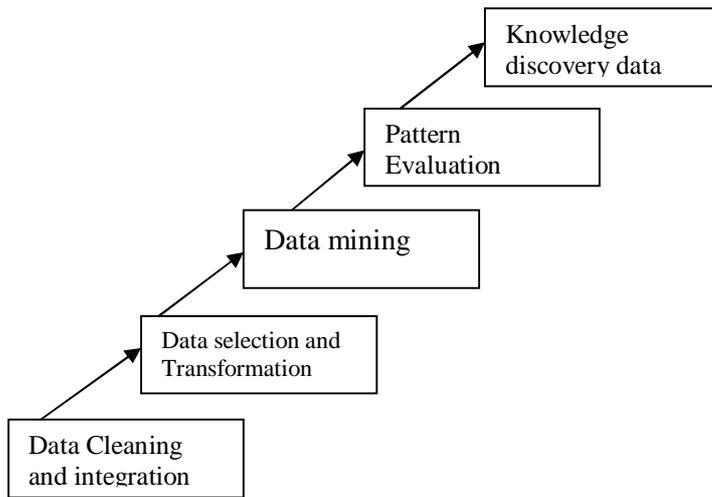

Figure 1 . The steps of extracting knowledge from data

Association Analysis:

This area of data mining aims at analyzing data to identify consolidated occurrence of events and uses the criteria of support and confidence. It is known to be applied in student behavior [7]. Mining association rules searching for interesting relationships among items in given data set. In our data set association rule mining is used to identify possible grade values i.e. Excellent, Good, Average, Poor, Fail.

[Attendance=poor, Assignmet=poor,GPA=poor]$\rightarrow$ [Grade=poor]
(Support: 0.196, confidence: 0.757)

[Attendance=poor,Sessional=poor,GPA=poor]$\rightarrow$ [Grade=poor]
(Support: 0.166, confidence: 0.657)

[Assignmet=poor,Final_grade=poor,GPA=poor]$\rightarrow$ [Grade=poor]
(Support: 0.176, confidence: 0.737)

[Attendance=poor,Sessional=poor,Final_grade=poor]$\rightarrow$ [Grade=poor]
(Support: 0.296, confidence: 0.747)

The resulting Association rule depicts a sample of discovered rules from data for student with poor grade along with their support and confidence. To interpret the rules in association rules model, the first rule means that of engineering students under study, 19%(support) are poor in attendance, poor in assignment, having poor in GPA. There is 75% probability or confidence that student will get the grade poor and so on.

Classification:

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups**.** A Rule-based classification extracts a set of rules that show relationships between attributes of the data set and the class label. It used a set of IF-THEN rules for classification.

If Attendance=excellent and Assignment=good and Sessional marks=excellent and GPA=good and Final_grade=excellent, then excellent.

If Attendance=excellent and Assignment=good and Sessional marks=good and GPA=good and Final_grade=good, then good.

If Attendance=average and Assignment=good and Sessional marks=good and GPA=good and Final_grade=good, then average.

If Attendance=poor and Assignment=poor and Sessional marks=average and GPA=poor and Final_grade=poor, then poor.

Association rules are characteristic rules (it describes current situation), but classification rules are prediction rules for describing future situation.

Clustering:

Clustering is a division of data into groups of similar objects. From a machine learning perspective clusters correspond to hidden patterns, the search for clusters is unsupervised learning, and the resulting system represents a data concept. From a practical perspective clustering plays an outstanding role in data mining applications such as scientific data exploration, information retrieval and text mining, spatial database applications, Web analysis, CRM, marketing, medical diagnostics, computational biology, and many others[8]. The K-means algorithm, probably the best one of the clustering algorithms proposed, is based on a very simple idea: Given a set of initial clusters, assign each point to one of them, and then each cluster center is replaced by the mean point on the respective cluster. These two simple steps are repeated until convergence [9]. The objective of this k-means test is to choose the best cluster center to be the centric. The k-means algorithm requires the change of nominal attributes in to numerical. The clustering method produced a model with five clusters.

ATTRIBUTE	CLUSTER1	CLUSTER2	CLUSTER3	CLUSTER4	CLUSTER5
ATTENDANCE	73.088	67.677	65.433	60.878	58.566
ASSIGNMENT	70.097	67.675	65.986	57.657	54.233
SESSIONAL	68.975	64.886	62.984	57.564	53.568
GPA	11.717	10.787	9.877	7.986	6.764
FINAL GRADE	84.055	81.785	77.886	70.764	63.986

V. CONCLUSION

In the present study, we have discussed the various data mining techniques which can support education system via generating strategic information.. Since the application of data mining brings a lot of advantages in higher learning institution, it is recommended to apply these techniques in the areas like optimization of resources, prediction of retainment of faculties in the university, to find the gap between the number of candidates applied for the post, number of applicants responded, number of applicants appeared, selected and finally joined. Hopefully these areas of application will be discussed in

REFERENCES

[1] C. Romero, S. Ventura, E. Garcia, "Data mining in course management systems: Moodle case study and tutorial", Computers & Education,Vol. 51, No. 1, pp. 368-384, 2008

[2] C. Romero, S. Ventura "Educational data Mining: A Survey from 1995 to 2005", Expert Systems with Applications (33), pp. 135-146, 2007

[3] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, **"**Data Mining Model for Higher Education System", Europen Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010

[4] K. H. Rashan, Anushka Peiris, "Data Mining Applications in the Education Sector", MSIT, Carnegie Mellon University, retrieved on 28/01/2011

[5] Han Jiawei, Micheline Kamber, Data Mining: Concepts and Technique. Morgan Kaufmann Publishers,2000

[6] W.M.R. Tissera, R.I. Athauda, H. C. Fernando "Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining", IEEE International Conference on Information Acquisition, 2006

[7] Sri Lanka Institute of Information Technology, *http://www.sliit.lk/* , retrieved on 28/02/2011

[8] Sun Hongjie, "Research on Student Learning Result System based on Data Mining", IJCSNS International Journal of Computer Science and Network Security, Vol.10, No. 4, April 2010

[9] Academy Connection – Training Resources In html, http://www.cisco.com/web/learning/netacad/index, December 28th, 2005.

[10] Wayne Smith, "Applying Data Mining to Scheduling Courses at a University", Communications of the Association for Information Systems, Vol. 16, Article 23, 2005

[11] Firdhous, M. F. M. (2010). Automating Legal Research through Data Mining. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 9-16.

[12] Jadhav, R. J. (2011). Churn Prediction in Telecommunication Using Data Mining Technology. International Journal of Advanced Computer Science and Applications - IJACSA, 2(2), 17-19.