1. INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. It sits at the common frontiers of several fields including Data Base Management, Artificial Intelligence, Machine Learning, Pattern Recognition, and Data Visualization. From a statistical perspective it can be viewed as computer automated exploratory data analysis of (usually) large complex data sets. In spite of the somewhat exaggerated hype, this field is having a major impact in education, business, industry, and science. It also affords enormous research opportunities for new methodological developments. Despite the obvious connections between data mining and statistical/mathematical data analysis, most of the techniques used in Data Mining have so far originated from the field of Statistics. In this research, we will try to explore some of the new models and techniques used to higher education systems.

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery.
1.1 Overview of Data Mining Education System

Severe challenges are being faced by students and alumni in higher education. 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? What groups of alumni are most likely to offer pledges? In addition to these challenges, 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 study addresses the capabilities of data mining and its applications in higher education.

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.

In order to produce new observations from existing data there is a need to have a 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.

1.3 Motivation

It was very early recognized by people in various departments of organizations—for example sales and marketing—that their company can become much more competitive in a constantly changing business environment if they could get hold of information regarding their customers’ shopping habits. If the people in charge of promoting the company’s products, for example, had this information, they would be able to apply direct marketing techniques to every customer. In other words, it would have been possible to target specific customers so that no money would be wasted in advertising campaigns and the content and layout of the company’s web page, if there was one, could be altered to comply with the observations made.

This, however, became possible only quite recently when advances in computer science, such as the invention of smaller and much more powerful processors as well as the parallel processing, made possible the collection and analysis of such data. This, coupled with the unprecedented growth of the World Wide Web as a communication medium, lead to the growth of this scientific field as it was attracting the attention of many big organizations and universities.

The educational industry, however, has not been given too much attention with respect to this area. In literature and on the World Wide Web we can find many examples of companies that claim to have enjoyed great benefits from employing data mining techniques; on the other hand it will be very hard to find a university’s case study, if any.

1.4 Problem Statement

A very promising area for attaining this objective is the use of data mining. Data mining or knowledge discovery in databases (KDD) is the automatic extraction of implicit and
interesting patterns from large data collections. Next to statistics and data visualization, there
are many data mining techniques for analyzing the data. Some of the most useful data mining
tasks and methods are clustering, classification and association rule mining. These methods
uncover new, interesting and useful knowledge based on users' usage data. In the last few
years, researchers have begun to apply data mining methods to help instructors and
administrators to improve educational systems.

Association rule mining is one of the most well studied data mining tasks. It discovers
relationships among attributes in databases, producing if-then statements concerning
attribute-values. Association rule mining has been applied to web-based education systems
from two points of view: 1) help professors to obtain detailed feedback of the e-learning
process: e.g., finding out how the students learn on the web, to evaluate the students based on
their navigation patterns, to classify the students into groups, to restructure the contents of the
web site to personalize the courses; and 2) help students in their interaction with the e-
learning system: e.g., adaptation of the course according to the apprentice's progress, e.g., by
recommending to them personalized learning paths based on the previous experiences of
other similar students. Therefore, following research questions needs to be explored:

- How can data mining help find students data?
- How can data mining help to group student's behavior, and solve their problems?
- How can some associative rules be found between students that could cluster their types of problem?
- How can data mining be used to identify those students who are at risk especially in very large classes?
- How can data mining be used in student's fees details and subjects?
- How can data mining be used to grade student's academic progress
1.5 Objective of the work

The field of Data Mining is concerned with finding new patterns in large amounts of data. Widely used in Business, it has scarce applications to Education. Here we are interested in mining student models in a pedagogical perspective. The objective of our study is to address how to make data possible to mine, to identify which data mining techniques are useful and understand how to discover and present patterns that are pedagogically interesting both for learners and teachers.

In the association rule mining area, most of the research efforts went in the first place to improving the algorithmic performance and in the second place into reducing the output set by allowing the possibility to express constraints on the desired results. Over the past decade a variety of algorithms that address these issues through the refinement of search strategies, pruning techniques and data structures have been developed. While most algorithms focus on the explicit discovery of all rules that satisfy minimal support and confidence constraints for a given dataset, increasing consideration is being given to specialized algorithms that attempt to improve processing time or facilitate user interpretation by reducing the result set size and by incorporating domain knowledge.

Most of the current data mining tools are too complex for educators to use and their features go well beyond the scope of what an educator might require. As a result, the course administrator is more likely to apply data mining techniques in order to produce reports for instructors who then use these reports to make decisions about how to improve the student’s learning and the online courses. Nowadays, normally, data mining tools are designed more for power and flexibility than for simplicity. There are also other specific problems related to the application of association rule mining to e-learning data.

Association rule mining algorithms need to be configured before they are executed. So, the user has to give appropriate values for the parameters in advance (often leading to too many or too few rules) in order to obtain a good number of rules. Most of these algorithms require
the user to set two thresholds, the minimal support and the minimal confidence, and then find all the rules that exceed the thresholds specified by the user. Therefore, the user must possess a certain amount of expertise in order to find the right settings for support and confidence to obtain the best rules.

Educational data sets are normally very small if we compare them with databases used in other data mining fields — typical sizes are the size of one class, which are often only 50-100 examples. In very few cases, we get data from 200-300 students. Therefore, the application of traditional association algorithms will be simple and efficient. However, association rule mining algorithms normally discover a huge quantity of rules and do not guarantee that all the rules found are relevant.

Event though support and confidence already allow for pruning many associations, often it is desirable to apply other constraints as well; for example, on the attributes that must or cannot be present in the antecedent or consequent of the discovered rules. One aspect that is often overlooked when applying data mining techniques on small datasets is that of overfitting. When there is only limited data available, and the hypothesis space (i.e., the total number of possible rules) is huge, it is inevitable in a statistical sense, to get false discoveries because every rule has a small chance of being true in the given data, and the number of rules is huge. To address this problem it is more than advisable to restrict the hypothesis space by inserting as much as possible background knowledge; i.e. restricting the body and heads of the rules as much as possible, and maybe, in worst case, even remove numerous attributes from the dataset. When evaluating the outcome of an association rule mining operation, one also has to take very well into account that without further controlled experiments the found associations are, at least, open to debate.

The objectives of this study are: to obtain an optimal predictive model for students within such systems which could help students use the learning resources better, based on the usage
of the resource by other students in their groups, help teachers design their curricula more effectively and provide the information that can be usefully applied by teachers to increase student learning, student records such as admission data, admitted enquiry, fees details, subjects details, performance in every subjects and evaluation for every students. Another objective is to find the best technique for extracting useful information from large amount of data in an educational system.

1.6 Research Approach and Methodology
To carry out the proposed research work following methodology is proposed:

1.6.1 Orientation
The research starts with the orientation on the area of education data mining, what are data mining, education data mining about and which education system issues are in dire need of investigation by consulting websites of current university education system service offerings, reading news articles and discussing data mining application for education system and online education system with professionals.

1.6.2 Data collection
In this phase data are collected, mainly data related to basic concept of data mining, online education system in data mining, essential characteristics, architecture, secure educational key management model, benefit, advantages and disadvantages, client and server architecture, how to configure online education system, challenges and education system goals etc.

1.6.3 Data Analysis and Experiment
Once the data are collected analysis will be undertaken. During this phase the performance of education data mining system is analyzed which includes various techniques and also secure education key management model in university education system.
1.6.4 Organization of the Thesis

This thesis is organized as follows:

First chapter introduction, different definition in data mining, provides the overview of data mining educational systems, motivation behind the study, research issues and problem statements and the objectives of this thesis. It lays the foundation for the whole thesis.

Second chapter data mining characteristics, architecture, current data mining in education system and various related concepts and methodologies which address the related work in education Data Mining system and some application scenarios.

Third chapter presents analysis of performance of education model using data mining techniques.

In chapter four proposed educational model is tested using experimental analysis of data mining techniques. For experimental analysis, educational data sets are used and various aspects are also discussed to provide recommendations for promoting an educational model in data mining. This chapter discusses the methods to find interesting association rules within the student’s databases. We propose a framework for the discovery of interesting association rules within a educational system.

In chapter five presents the implementation and results of different classification and clustering modeling techniques on the Higher Education data. This introduces the basic concepts of classification and presents methods for evaluating and comparing the performance of a classification technique. While its focuses mainly on a technique known as decision tree and this chapter is also applicable to other classification techniques as well as clustering methodology’s theory, implementation, and results.

Chapter five concludes the study and suggests the future work related to this study.

An exhaustive list of more than 140 references in the area of imputation techniques and model in education Data mining.