



Munich Personal RePEc Archive

# **Role of Educational Data Mining Model in university course selection by the students.**

Rehman, Mobeen

COMSATS Institute of Information Technology, Virtual Campus,  
Islamabad, Pakistan

January 2015

Online at <https://mpra.ub.uni-muenchen.de/62680/>  
MPRA Paper No. 62680, posted 17 Mar 2015 05:04 UTC

**Role of Educational Data Mining Model in university course selection by  
the students.**

First Draft  
January, 2015.

Mobeen Ur Rehman  
COMSATS Institute of Information Technology, Virtual Campus  
Islamabad, Pakistan  
Mobeen@vcomsats.edu.pk

# **1. Review of Literature**

## **1.1 Data Mining**

There are various definitions of database in different contexts by the previous researchers. According to Matheus et al (1991), data mining is a process to discover unobvious but useful information. Curt (1995) defined data mining as a process of transforming vocabulary of unorganized nature to an organized data for some useful purpose. This organized data can be used as a refined knowledge from which a decision tree can be constructed. Smyth et al (1996) described data mining as a simple process that results in a discovery of comprehensive, valid and useful data. According to Hui and Jha (2000), data mining is an analysis of semi-automation and automation so that the meaningful rules and relationships can be discovered out of a large set of data in a database system. Adams et al (2000) suggested that data mining is a process of discovering valuable and interesting information of useful nature from the existing database. According to Han and Kamber (2006), data mining is a process that involves the discovery of hidden images, knowledge and patterns from large sum of data to forecast or predict behaviors. Gartner (2007) describes data mining as the process of finding correlation of meaningful value, trends and patterns from large amount of data stored in the repositories through mathematical and statistical techniques and technologies involving pattern recognition.

## **1.2 Data Mining Role is Knowledge Discovery**

Knowledge discovery in data base refers to the process of transforming, analyzing, enduring, clarifying and conforming data within the customer already present in the system or an outside one. This finally results in the formation of backup decision making system with continuous maintenance and improvement. Peacock (1998) further subdivided data mining in narrow and a broad category. The narrow category has its limitation in the methodology regarding mechanical learning emphasizing the discovery processes using artificial intelligence i.e., neural networks, decision tree algorithms, correlation rule and genetic algorithms. On the other hand, broad category of data mining stresses the importance of knowledge discovery in database.

## **1.3 Advantages of Data Mining**

One of the major advantages of data mining according to Thearling et al (2001) is its ability to forecast rather than relying merely on tracing back the existing data. Hui and Jha (2000) recommended that the data mining process should include establishing the mining objectives using domain knowledge to get the relevant data in accordance with the established goals and objectives. Further selection of data by identifying the variables characteristics is of utmost importance so that the useful mining can be performed. After that data pre-processing and data transformation in a new format are also important. Data warehousing (envisioning, building, planning) by enhancing and managing database along with data mining mainly through correlation structures the concluding stages of the overall process. Evaluation of the results includes elaboration by undergoing the above mentioned comprehensive process.

#### **1.4 Data Mining in Education Sector**

Luan (2004) comprehends that the data mining techniques uncover the patterns that are hidden in the massive volume of data by using the reporting capabilities of the organizations. This techniques needs not only to be applied in the education sectors rather it has its applications in banking sector, fraud detection, biomedical, retail and marketing industry etc. (Luan 2002). In recent days, many of the technical education sectors are experiencing a large influx of data of structured and unstructured nature through intakes of students. This results in many issues surfacing from the knowledge gap that is created in such situations and hinders the achievement of quality goals and objectives.

#### **1.5 Need for Data Mining in Education Sector**

In the context of designing better tools to gauge student related issues, universities and educational institutions needs better assessment techniques and tools for analysis and making forecasts. This useful extracted information in the university environments like students information, class schedule, teachers information, alumni information, course contents and online information can facilitate many tasks like selection of better course for the students, performance, managing students, dropouts etc. so that student counseling, registration and student evaluation tasks can be performed in a better way. Making use of the data mining technique in improving the quality of graduate students is suggested by Waiyamai (2003) and he used ARM and cluster classification techniques so that patterns of previous students with major courses can be observed.

## **1.6 Designing Educational Data Mining**

The applications of data mining technique in education sector for the purpose of knowledge discovery is quite significant, not only to the organizations itself but to the students as well. Decision support system of the educational sector is supported by the data driven by knowledge. Educational data mining helps in learning enhancement through various trends in education like performance of students in various programs, selection of courses, trainings and the faculty development programs. By making use of the regression analysis, many of the factors are related to the performance of the students like family income and parent education etc. (Pal and Bhardwaj 2011). Many of the objectives can be achieved through the application of data mining techniques like retention rate of the students, improvement in educational ratio and increased learning outcome of the students. Therefore many of the hidden relationship and patterns can be revealed through the application of data mining on large data volume for the purpose of effective decision making (Ojha and Bala 2012). The process of data mining combines the process of machine learning, visualization techniques and statistics in discovery and knowledge extraction. Techniques like neural networks and decision tree are used to analyze the collected data through feedback forms and questionnaire so that the student's response towards employed university technologies, trends or educational pattern and instruction methodologies can be gauged. Different data mining models like decision trees, support vector machines, Naive Bayes, Linear regression, K-means, Minimum description length and O-Cluster can help in tapping students behavior pattern, retention rate, personalized intervention strategy and the prediction of course suitability (Zhang and Oussena 2010).

## **1.7 Various Existing Techniques for Data Mining**

The development of the data mining software should be in such a way so that it allows its user to make data analysis in different dimensions thereby enabling it and making summary of the results (Han and Kamber 2000). The data mining techniques are not limited to the traditional education institutions rather it can be applied on the virtual or distant educational system as well. According to Bhardwaj and Pal (2011) many tools like DBMiner, Intelligent Miner, Weka, filtering and visualizations etc. can perform data mining algorithms. Waiyamai (2003) found out that the patterns of the students who were likely to be good performers in the given subjects. The use of classification and clustering analysis can help in studying the relationship between learning styles of the students and their performance. He also identified students with good programming and explicit knowledge learning behavior through individual characteristic patterns. Haddawy et al (2007) also used many techniques like decision tree, prediction and Bayesian network etc. to check the suitability of data mining in predicting academic performance of the students. They conducted two case studies in finding the patterns of performance of graduate and undergraduate students. According to Oyelade et al (2010), performance of various graduate students according to their previously achieved scores by using K means clustering can help in making analysis of the academic performance of students in higher education institutions. Quadri et al (2010) analyzed the reasons of dropout of students in educational institutions by using decision tree analysis. In his study, Kovacic (2010) studied in detail the enrollment data of students to identify the most important factor of success in completing the registered

courses. He used statistical techniques like CART (Classification and regression tree), CHAID exhaustive, and QUEST classification tree methods. Ramaswami and Bhaskaran (2010) further used CHAID prediction model (based on classification tree Algorithm) to find the factors responsible for slow academic performance of slow learners.

### **1.8 Data Mining in Students Course Selection**

According to Librizzi et al (2004) many universities perform the process of evaluation of courses and programs including such courses through the end of semester ad hoc evaluation forms. The purpose of such evaluation forms and feedback forms is to make the quality of offered courses better and to devise strategies in the content formulation and course/faculty improvements. This process helps in improving the overall quality of faculty learning, courses and teaching methodology. Although the above mention process offer ways to make improvement in the overall course design, two factors can pose some issues is such response collection. First one is the way in which the items are selected and arranged and second is the presence of the potential confounding factor that can substantially influence the student's perception regarding quality of university courses (Gittoes and Draper 2004).

For the comparison among heterogeneous groups, characteristics of the group level and individual level should be adjusted to avoid misleading conclusions (Hand and Fayers 1997). In recent years many of the studies have been conducted to explore the

advantages and weaknesses of mixed effect models in handling the latent variables especially when the hierarchical structure is involved (Wu et al 1997).

### **1.9 Instrument Development for Data Mining**

Recent development of the Item Response Theory (IRT) serves as the probabilistic framework to deal with the likert scale of multi item response. Such instrument has the advantage of making use of the information contained in the response provided by the students through interpreting the characteristics of the included items in the questionnaire and the composition of the overall evaluation form of the student. According to Hoyt and Brown (1999), techniques like (CIRP) Institutional research program Freshman Survey and ASQ (Admitted Student Questionnaire) can be used to collect information from the students so that a complete academic research regarding the choice of academic program selection can be conducted. According to Connor et al (1999), the development of the checklist to tap the responses of the students is critical to this decision making by the students.

According to Bradley et al (2005), educational evaluation process must include the included course aims, outcomes and objectives. The perceptions of the students in their selection criteria of the courses also plays an important role as this helps not only in the course selection but also in the evaluation of the taught courses at the end of the semester. Many universities have made these criteria as the mandatory one rather than the optional one not only for course evaluation but for the teaching methodology improvement as well. Many situations in various universities ensure the involvement of such evaluations for the performance of the relevant staff along with the

improvement in the included courses. However, there may be some debate regarding the nature of the included courses i.e. compulsory or elective programs.

Bassin's study used the likert scale approach and followed a narrow style of evaluation by involving the five aspects of teaching namely lecture quality, text suitability, exam quality, consideration and participation. Drago and Wagner (2004) concentrated on the aspects like relation between preferred learning style and teaching effectiveness rather on the type of offered courses. The above discussion also confirms Collins (1996) findings about the academic disciplines by equally distributing participants fairly and widely across all disciplines. According to Emerson et al (2000), the method of evaluation by the students is a recognized one and it has been in practice for quite a long time. One important point to note is that the rating received by the student in accessing the quality of lectures not only includes factors like clarity of methodology, lecture class, readings etc. but also on the personal traits and expectations of each student. Therefore taking the mean and fair value of these collected responses is important to adjust for such biases. Many of the multilevel models can serve this purpose (Goldstein 2013).

### **1.10 Extension in traditional models**

Many of the extensions can be made in the traditional models like the consideration of IRT models into generalized nonlinear or linear mixed models. This can help in handling the nested units in clusters at higher level, covariates as a source of heterogeneity in the ratings of the students and finally in the latent traits of multi dimensions.

### **1.11 Factors Affecting Choice of Courses**

There are a lot of choices having an effect on the courses choice of high school students. If these factors are to be identified, there can be much help for industries, parents and educators to focus on these key factors during the selection of the career process. Along with these institutions, students will also benefit from this process for selecting suitable career for themselves. Many of the choice factors have been identified in the past researches with huge variations. Such variation suggest that there are a lot of factors that can shape the students preference regarding the selection of college programs and the courses in these programs (Hoyt and Brown 1999). According to Nora & Cabrera (1992), research has shown that the factors having an influential role changes from early educational life to later education level, aspirations of the career, ability, universities attributes, socio-economic status, encouragement from the parents and financial limitations all have important considerations on the college choice. According to Kinzie et al (2004), there are more main factors that have an influential role on selection of the college, which were the same in 1965 namely intellectual emphasis, advice of others, practicality and social emphasis.

Applicant's characteristics, universities perception, information sources, financial aid and competition among the institutions can be the obtained information that can further be refined through the application of principal component analysis after detailed interrogation with the determination of expectancy value (Braxton et al., 1995). According to Connor et al. (1999) the choice process has been found to widen the student participation as well as participation rates. Many factors like cost, access,

school and life experience, information and academic achievement have influencing power for institutional choice which is governed by rationality, multi factors and pragmatism (Moogan and Baron 2003). According to Foskett et al (2003), the role of perceptions at the young age is also influential in the choice decision and the variance in the factors involved in such decision making is quite variant along with their perceived importance.

Christe et al (2001) suggests that many of the students underestimate the cost of the higher education and therefore such costs acts as a barrier for them. Many of the students do not have knowledge right from the beginning about the university characteristics whereas some students face trouble due to difficulty of the available information that deters them from applying in the universities (Forsyth and Furlong 2003). According to Archer & Hutchings (2000), support from schools and families of the candidates play a crucial role whereas many other researchers suggest the some non-traditional students have different experiences regarding the choice process (Connor et al. 2001).

According to Wood and Lange (1998), reputation of the universities along with its employment record is the major factor in the selection of student regarding the universities. Connor et al (1999) suggests that the choice of the right subjects that needed to be included in the curriculum is quite influential. Along with the choice of the offered subjects, support from the university's academic department helps the students.

According to Connor et al (1999), besides the choice of right course, many other factors like employment prospects, image, teaching requirements, academic support

locations and facilities are included as the most influential for the applicant of universities. Therefore, there is not any room for the choice rather a strong need for the development of checklist is required for students help in their decision's choice. According to Stroade et al (2001), the choice of students also depends upon the selection of elective or compulsory courses as it represents a variable relating to the students attitude towards a particular educational program. Likert scale finding of Gage (1961) suggested that students of elective courses grade their instructors higher as compared to the teachers of the required courses. Same finding has also been confirmed by Lovell and Haner (1955). It was also found that the instructors of the non-quantitative courses received higher ratings by the students as compared to the quantitative courses.

According to Splaver (1977), role of personality is also important in selecting the suitable career for oneself. This type of personality includes self-motivation in the early stages rather than the procrastinating behavior that requires waiting until compulsion. Further Splaver (1977) also mentioned that good understanding of oneself and one's personality is required to make intelligent decisions for the career plans. He also mentioned that the counselors do not completely share the success stories of ones currently in their jobs as compared to the students who traditionally completed their education in a four year college time.

According to Victor, E.R. (2013), one of the reasons of under employment and unemployment is the lack of competence and skills of graduating students required by the companies and can be attributed to the wrong selection of courses led by unguided decision making at the very start of the college life. Clemena and Marie (2002) argued

that the selection of courses by majority of the students is made not on the basis of interest or liking but because of the better prospects regarding employment. Students should be quite capable of selecting and evaluating the courses included in their degree program so to develop and formulate long term career goals and educational plans (Clemena and Marie 2002).

According to Pascual (2014), most of the students make choices as per their interest and skills about the courses in the third semester of their degrees. This selection of courses in the later stage of their degrees also represents their future preferences at the post-graduation university levels whereas according to Pastorelli et al (2001), many other factors like intellectual abilities and aptitude of the students play an important role along with the personality of the students and their interest level.