JOURNAL OF COMPUTING & ORGANISATIONAL DYNAMICS
Double Blind Peer Reviewed Open Access Journal

Mining Postgraduate Students' Data Using Apriori Algorithm

Yousef Ibrahim Abu Zawayda (yzwayda@gmail.com)

Faculty of Information Technology

Universiti Utara Malaysia

ABSTRACT

Data Mining is more than just traditional data analysis. It uses traditional analysis tools like statistics and graphics in addition those associated with artificial intelligence such as rule induction. It is a characteristic approach or attitude to data analysis. The emphasis is not so much on extracting facts, but on generating hypotheses. The aim is more to yield questions rather than answers. Insights gained by data mining can then be confirmed by traditional analysis. The discussion made in the previous section highlighted several problems and opportunities. Firstly, as there is an abundance of information contained in UUM's postgraduate database, useful knowledge extracted from the data could benefit the university. Secondly, the capability of Apriori algorithm to mine students' data has not been explored. This study could provide evidence on the capability of Apriori in mining data. Thirdly, there is a need to identify factors that can influence students on deciding to enroll in postgraduate programs. Identifying these factors could benefit UUM as this information could direct the management to run promotion upon any groups of potential students. In this study, the Knowledge Discovery in Databases (KDD) process methodology was implemented. The study met the set objectives of the study which are clearly discussed in the findings of the study.

Keywords: Data mining, Webmining , Algorithm, Apriori

Print ISSN 2306-9309 Online ISSN: 2306-9295

Introduction

The word mining has been used to describe the activity of digging coal or other essential substances out of the ground. The Cambridge Advanced Learner's Dictionary defines the word mining in several ways. As an information technology jargon, mining commonly implies data mining, and is defined as applying a specific algorithm for the discovery of hidden knowledge, unexpected patterns and new rules in large databases, (Dunham, 2003). Fayyad *et al.* (1996) defined data mining as "the use of algorithms to extract the information and patterns derived by the Knowledge Discovery Database (KDD) process". In short, data mining is the process of analyzing data from different perspectives and summarizing the results as useful information.

Data Mining is more than just traditional data analysis. It uses traditional analysis tools like statistics and graphics in addition those associated with artificial intelligence such as rule induction. It is a characteristic approach or attitude to data analysis. The emphasis is not so much on extracting facts, but on generating hypotheses. The aim is more to yield questions rather than answers. Insights gained by data mining can then be confirmed by traditional analysis. Additionally, data mining offers a solution: automatic rule extraction. By searching through large amounts of data, one hopes to find enough requests of an association between data value occurrences to suggest a statistically significant rule. However, a domain expert is still needed to conduct and evaluate the process and to apply the results.

Data mining techniques have been widely applied to solve problems in the industry. Lau and Gao (2005) proposed a data mining approach for performance management in the banking industry. Meanwhile, in science Shi and Jaja (2002) considered the problem of organizing large scale earth science raster data to efficiently handle queries for identifying regions whose parameters fall within certain range values specified by the queries. Banks *et al.* (2004) have developed an undergraduate data mining course that could be taught in semester or quarter systems and within institutions of varying demographics. They also established methods for an approach to increase student retention in such a course, and identified data mining skills essential for problem solving. In engineering, data mining was used to diagnose faults of boilers in thermal power plants. Yang and Liu (2004) proposed a hybrid-intelligence data mining technique to extract hidden diagnosis information from Supervisory Control And Data Acquisition (SCADA) system. Ambwani, (2003) applied multi class support vector machine classifiers using one-versus-one method on *government* data. The aim was to identify attack precisely by type. In business, Shang *et al*. (2004) designed a novel mining system known as WebCom Miner to help a company to scientifically make decisions for their products.

In the field of education, several researches have been found. Among these researchers are Ha *et al.* (2000), Minaei-Bidgoli *et al.* (2003), Holt and Chung (2002), Watanabe and Nakayama (2003), Zhou (2003), Zang and Lin (2003), Chan *et al*. (2003), and Fernandez *et al*. (2001). Further details on their works are discussed in the literature review.Hence, this study apply one of the data mining techniques namely Apriori Algorithm in the field of education. The aim is to mine the postgraduate data of Universiti Utara Malaysia (UUM) to scale for pattern of enrollment.

The *Apriori* algorithm searches for large itemsets during its initial database pass and uses its result as the seed for discovering other large datasets during subsequent passes. Rules having a support level above the minimum are called large or frequent itemsets and those below are called small itemsets (Chen *et al*., 1996). However, Apriori algorithm has been found to be a good technique in data mining and has been applied in various areas such engineering (Zaiane *et al*., 2001), business (Chan *et al*., 2003), medical (Doddi *et al*., 2002), and marketing (Brin *et al*., 1997).

Regarding the issue on education, recently, UUM announced that increasing postgraduate enrolment rates is a top priority. Figure 2 shows example of the graph of postgraduate intake, which in UUM since 1997.

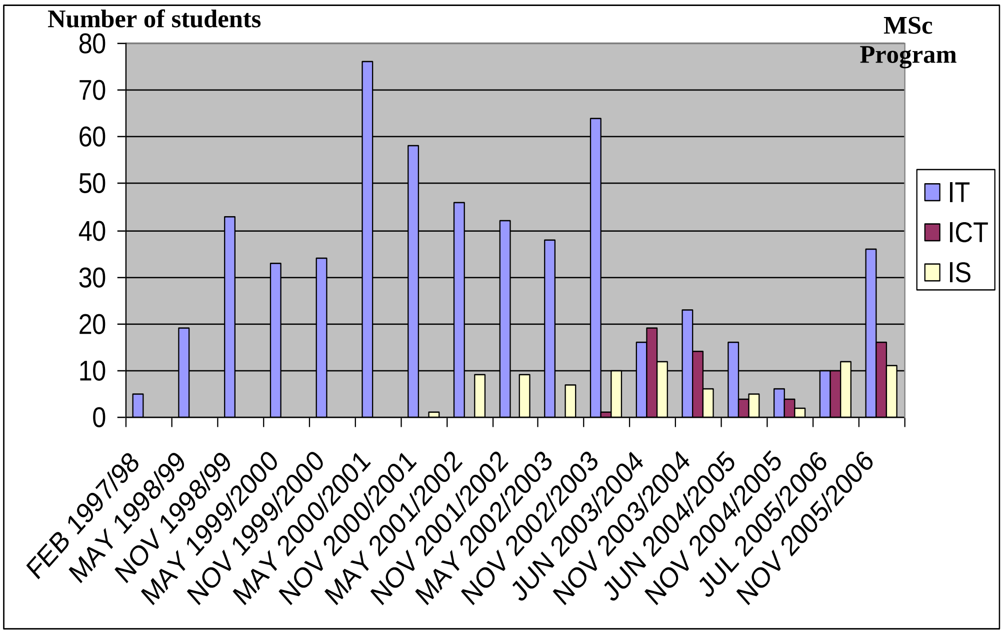

Figure 2: The number of students enrolling MSc programs

Print ISSN 2306-9309 Online ISSN: 2306-9295

The graph shows that in the late 90s, an increase of intake was observed, which peaked in 2000/2001. However, after 2001, the intake slowly decreased and later increased again in 2002/2003. After 2003 it can be seen that there is no substantial increase in intake as what was experienced from year 1998 till 2001. Thus, the slow growth of intake after 2003 has resulted the need of UUM's top management to come up with a good strategy/program to improve the situation. One way was to set up good promotional activities. All the faculties are requested to set up a number of promotional strategies for attracting more postgraduates, including international students.

Thus, in planning the strategies, the management definitely requires some knowledge on who will be likely to enroll in postgraduate programs offered by UUM. One of the solutions is to analyze the type of students that have undertaken UUM postgraduate programs. Thus, due to this, the postgraduate students' data is chosen to be mined. Another reason is that, the abundance of data contained in UUM postgraduate students database may lie some hidden extra information that can serve as useful knowledge. The knowledge if identified could direct the management to run strategic promotions for capturing more postgraduate students. The knowledge could also be used to design the structure of academic programs that can improve teaching and learning process. Thus, this can indirectly improve the quality of the offered program.

Results from the past researches highlighted above indicated that the use of Apriori produced satisfactory results on the whole. Thus, for this reason Apriori algorithm is chosen to mine students' data for finding relationships between factors. Specifically, this study aims to mine students' data, namely postgraduate students data based on several reasons: (i) too much data but too little information, thus there is a need to extract useful information from the data and to interpret the data; and (ii) the appropriate AI technique of mining (exploring) the postgraduate students' database (for promotion segmentation) is not yet identified.

Problem Statement

The discussion made in the previous section highlighted several problems and opportunities. Firstly, as there is an abundance of information contained in UUM's postgraduate database, useful knowledge extracted from the data could benefit the university. Secondly, the capability of Apriori algorithm to mine students' data has not been explored. This study could provide evidence on the capability of Apriori in mining data. Thirdly, there is a need to identify factors that can influence students on deciding to enroll in postgraduate programs. Identifying these factors could benefit UUM as this information could direct the management to run promotion upon any groups of potential students. In addition, the knowledge can also be used to improve the teaching and learning. That can indirectly improve the quality of the offered program.

Objective of the Study

The objectives of this study are:

1.3.1 To find useful patterns in postgraduate programs data
1.3.2 To explore the capability of Apriori algorithm to mine students data
1.3.3 To identify which factors can influence students on deciding to enroll in postgraduate programs.

Scope

The sampling data set for this study collected from questionnaires answered by students who are currently undertaking postgraduate programs that include MBA(Acct), MBA(General), MBA(Tourism), MHRM, MSc(Finance), MSc(IT), MSc(IS), MSc(ICT), MSc(Mgt), MSc(Edu Mgmt), MSc(Acct), MSc(Banking), MSc(Int Acct), MSc(Mgmt) and MSc(Mgt). The study only concentrates to the use of Apriori Algorithm upon postgraduate data, the effectiveness of the algorithm has not yet considered.

Data Mining

Data mining computerizes the process of finding relationships and patterns in data and distributes results that can be also used in an automated decision support system or assessed by a human analyst. With the proliferation of data warehouses, data mining tools are widely used on the market. Their objective is to explore hidden knowledge in data. The term "data mining" use by many traditional reports and query tools and statistical analysis systems in their product descriptions.

There are several definitions for data mining, the field of data mining and knowledge discovery is emerging as a new, fundamental research area with important applications to education, science, business, engineering and medicine (Grossman *et al.*, 1998).

In other hand, data mining is also defined as the use of numerical analysis, mathematical techniques or visualization to identify non-trivial numerical relationships within a dataset to derive a better understanding of the data and to predict future results (Rygielski *et al.*, 2002).There's a lot of domain that used data mining applications in their platform such as in Business, Accounting, Finance, Credit Card, Healthcare, Engineering, Education, and many more.

Print ISSN 2306-9309 Online ISSN: 2306-9295

Applications of data mining

From all the researches mentioned above, it can be seen that data mining have a significant impact on industries. Through the use of various data mining tasks, it is believe that varieties of problems can be solved.

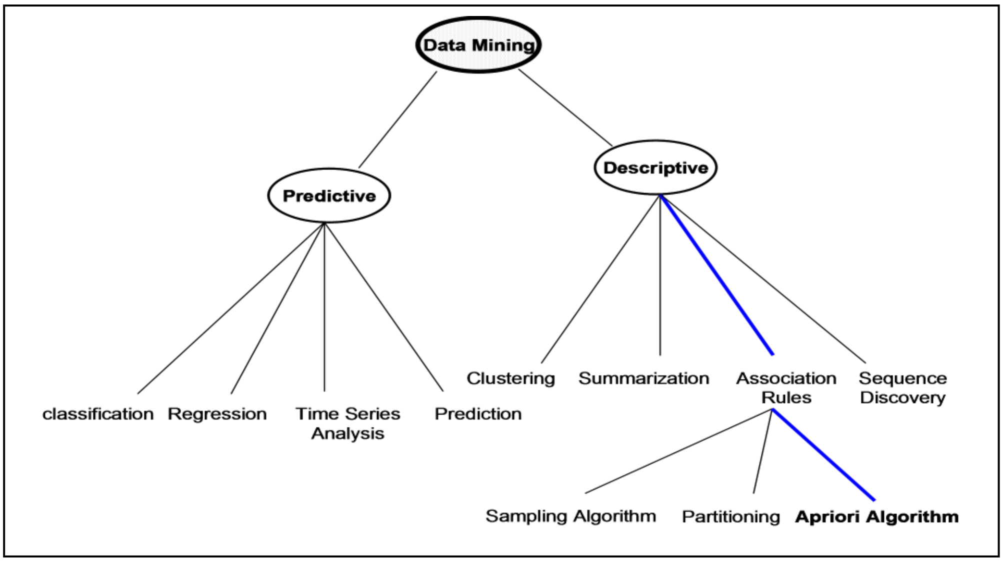

Figure 2.1: Data Mining Models and Tasks

As illustrated in Figure 1, data mining tasks can be classified into predictive or descriptive in nature. A predictive model makes a prediction about values of data using known results found from different data. Meanwhile, a descriptive model identifies patterns or relationships in data (Dunham, M. H., 2003). Unlike the predictive model, a descriptive model serves as a way to explore the properties of the data examined and not to predict new properties.

There are four approaches of descriptive data mining that include clustering, summarization, association rules and sequence discovery. This study will be concentrating on association rule approach that applies the Apriori algorithm.

Association rule is used to discover interesting associations between attributes contained in a database. This technique is also known as *market basket* analysis. Based on frequency counts of the number of items occur in the event, association rule tells if item X is a part of the event, then what is the percentage of item Y is also part of the event.

An association rule is a relationship of the form $X \Rightarrow Y$ where X is the antecedent item set and Y is the resultant itemset. An example of the rule can be, "customers who purchase an item X are very interested to purchase another item Y at the same time". There are primarily two measures of quality for every rule, support and confidence. The rule $X \Rightarrow Y$ has support s% in the transaction set N if s% of transactions in N contains $X \cup Y$. The rule has confidence c% if c% of transactions is N that contain X also contain Y. The objective of association rule mining is to find all the rules with support and confidence more than some user specified thresholds.

Several association algorithms are sampling algorithm, partitioning, and Apriori algorithm. Among the algorithm Apriori is the most common and most popular data mining technique for association based analysis (Agrawal and Srikant, 1995). The algorithm uses two values for rule construction: a support value and a confidence value. Depending on the setting of each index threshold, the search space can be reduced, or the candidate number of association rules can be increased.

Business/Accounting/Finance

Recently, we can see how data mining improves business process and this is the only way for getting good and efficient outcome, as opposed to purely manual and interactive data exploration.Businesses often need to forecast sales to make decisions about inventory, staffing levels, and pricing. Data mining have had huge success at sales forecasting, due to their capacity to simultaneously consider multiple variables such as market demand for a product, customer's disposable income, the price of the product, and the size of the population. Forecasting of sales in supermarkets and whole sales provisions has been studied by Thiesing *et al.* (1995) and the result have shown to perform well when compared to traditional statistical techniques like regression, and human experts.

The tools and technologies of data warehousing, data mining and other Customer Relationship Management (CRM) techniques proffer new opportunities for businesses to act on the concepts of relationship marketing. CRM is defined by four elements of a simple framework, which are *Know*, *Target*, *Sell* and *Service*. CRM requires the firm to know and understand its markets and customers. This involves detailed customer intelligence in order to select the most profitable customers and identify those no longer worth targeting (Rygielski *et al.*, 2002).

Another application in this area is the fraud detection. A variety of data mining techniques have been used to develop fraud systems which can discover fraudulent credit card transactions in near-real time. This problem is challenging due to the size of the data sets, the infrequency of the events of interest, and the performance

Print ISSN 2306-9309 Online ISSN: 2306-9295

requirements for near-real time detection. Data mining has also enhanced fraud detection in other application domains, including telecom fraud and insurance fraud (Grossman *et al.*, 1998).

In finance domain Blazejewski and Coggings (2004) employed histograms and self-organizing maps (SOM) to achieve unsupervised clustering and visualization of four dimensional trade level data for the ten stocks on the ASX with the largest market capitalization. They took note of that the SOM transformation give emphasis to regularly occurring trading patterns conditioned on all dimensions of the trade data, although the simple histograms fail to separate common and uncommon trading behaviors. This reveals that buyer-initiated and seller-initiated trades form two distinct clusters in correspondence with non-equilibrium market conditions and extracts the main structural features of the clusters. An innovative KDD approach was designed by Kauderer *et al.* (1999) to optimize the collection efforts in automobile financing. After 12 months in operation, their experience shows a steady performance of the model and enough robustness to allow for attains cut-off adjustments to address fluctuating workload and ability constraints. The applications of data mining in Business/Accounting and Finance can be summarized.

Table 2.1: Related studies on applications of data mining in Business, Accounting and Finance.

Business/Accounting/Finance		
Author	**Year**	**Title**
Blazejewski & Coggings	2004	Application of Self-Organizing Maps to Clustering of High-Frequency Financial Data
Rygielski, *et al.*	2002	Data mining techniques for customer relationship management
Bansal, *et al.*	2001	Neural Network Based Data Mining Applications for Medical Inventory Problems
Smith & Gupta	2000	Neural Networks in Business: Techniques and Applications for the Operations Researcher
Kauderer, *et al.*	1999	Optimization of collection efforts in automobile financing a KDD supported environment
Grossman, *et al.*	1998	A Report of three NSF Workshops on Mining Large, Massive, and Distributed Data
Scott, *et al.*	1997	Experiences of Using Data Mining in a Banking Application
Thiesing, *et al.*	1995	Short Term Prediction of Sales in Supermarkets

Print ISSN 2306-9309 Online ISSN: 2306-9295

Credit Card

According to Chan, *et al*. (1999), studied an important problem, which is scalable techniques to analyze massive amounts of transaction data that competently compute fraud detectors in a timely manner, especially for e-commerce. Besides scalability and competence, the fraud-detection task shows technical problems that consist of skewed distributions of training data and no consistent cost per error, both of which have not been widely studied in data mining and the knowledge discovery. They surveyed and evaluated a number of techniques that they have suggested and executed that address these three main issues concurrently. Their proposed methods of merging multiple learned fraud detectors under a "cost model" are general and apparently useful; their empirical results confirmed that they can significantly reduce loss due to fraud through delivered data mining of fraud models.

Medical

Medical is another field which requires and implemented data mining in order to increase its effectiveness and efficiency. Antonie *et al.*, (2001) performed a research on the application of data mining techniques for medical image classification for breast cancer. The research uses of different data mining techniques, neural networks and association rule mining, for anomaly detection and classification of breast cancer. The results showed that the two approaches performed well, they conducted demonstrate the use and effectiveness of association rule mining in image categorization.

Another research in medical is done by Walker *et al.*, (2004) used data mining to change the gene expression in Alzheimer brain. The research address the problem of dealing with microarray data that come from two known classes (Alzheimer and normal) and applied three separate techniques to discover genes associated with Alzheimer disease. Another application is in medical data (Doddi *et al.*, 2002), which their main objective is to explore the relationships between procedures performed on a patient and the reported diagnoses.

According to Tigrani (1998), data mining widely used in medical domain and also contributes in prostate cancer detection. Tigrani who is a data-mining analyst with in background in machine learning: hypothesis testing by the former and bagged classification models by the letter. The objectives of this application is to show the advances in data mining and machine learning research that have been motivated by medical data analysis. Furthermore, it analyse different and more complex tests, first hypothesizing possibly superior tests and using statistical hypothesis test to judge their merit, and then using data mining methods to explore a wide space of possible test.

According to Moser *et al***.** (1999) presented that Data Mining Surveillance System (DMSS) uses a large electronic health-care database from Birmingham Hospital for monitoring emerging infections and antimicrobial resistance. The information form DMSS can indicate potentially important shifts in infection and antimicrobial resistance patterns in the intensive care units of a single health-care facility. DMSS can automatically identify new, unexpected, and potentially interesting patterns in hospital infection control and public health surveillance data. Domain experts to determine their actual importance must investigate the events identified by DMSS. They have demonstrated that DMSS can identify potentially interesting and previously unknown patterns.

Table 2.2 shows several researches using data mining in medical domains.

Medical		
Author	**Year**	**Title**
Walker *et al.*	2004	Data Mining of Gene Expression Changes in Alzheimer Brain
Doddi *et al.*	2002	Discovery of Association Rules in Medical Data
Antonie *et al.*	2001	Application of Data Mining Technologies for Medical Image Classification
Moser *et al***.**	1999	Application of Data Mining to Intensive Care Unit Microbiologic Data
Tigrani	1998	Data Mining and Statistics in Medicine: An Application in Prostate Cancer Detection

Engineering

Several applications of data mining have been used in engineering, Astronomical Data application is one of them. Traditionally, the search for new galaxies, stars, and quasars has primarily been done by astronomers visually examining individual photographic plates. Classification algorithms from data mining have recently been used to automate this process yielding new astronomical discoveries. The classification algorithms are applied to derived attributes produced by image processing, such as the brightness, area, and morphology of sky objects. The approach has also proved useful for detecting new objects too faint to be observed by a manual analysis or traditional computational techniques. For the 2nd Palomar Observatory Sky Survey, this approach resulted in over a three-fold increase in the size of the catalog (Grossman *et. al.*, 1998).

Other application of data mining in engineering area is Agent Academy, a multi-agent development framework for constructing multi-agent systems, or single agents. The behavior and intelligence of each agent in the community can be obtained by performing data mining on available application data and the respected knowledge domain. We have developed Agent Academy, a software platform for the design,

Print ISSN 2306-9309 Online ISSN: 2306-9295

creation, and deployment of multi-agent systems, which combines the power of knowledge discovery algorithms with the versatility of agents. Using this platform, they showed how agents, equipped with a data-driven inference engine, can be dynamically and continuously trained. They also discussed a few prototype multi-agent systems developed with Agent Academy (Mitkas *et al.*, 2003).

Bariani *et al.* (1998) demonstrated an automatic knowledge discovery process from a database of images in a context of Automated Visual Inspection (AVI). AVI is one of the most explored application fields of computer vision for its importance in the quality inspection of industrial production lines, even under informal quality models. When modeling informal knowledge, one of the most critical point turns out to be the correct and efficient translation of human experience into a set of rule. The objective of this project was to focus on the use of machine learning in inspection of industrial work pieces. It showed how machine learning can be exploited for data mining purposes and more specifically and selecting a minimal set of visual primitives, in order to perform reliable and robust classification of the inspected component.

Approaches used in AVI tasks are often classified as quantitative and qualitative: a quantitative inspection has the goal of extracting measures of specific, well-known features of objects such as areas of parameters, starting from images; instead, qualitative inspection relies on decision process based on a set of visual information, generally learned by human expert: in these cases, this information is not quantitatively well defined, depending on the human training and are generally not stable enough for different context changes.

Table 2.3: Shows several researches using data mining in engineering domains.

Engineering		
Author	**Year**	**Title**
Mitkas *et al.*	2003	Application of Data Mining and Intelligent Agent Technologies to Concurrent Engineering
Bariani *et al.*	1998	Data Mining For Automated Visual Inspection
Grossman *et al.*	1998	A Report of three NSF Workshops on Mining Large, Massive, and Distributed Data

Print ISSN 2306-9309 Online ISSN: 2306-9295

Mining Education Data

In the field of education, data mining has been applied by many researchers to solve various problems. Among these researchers are Ha *et al.* (2000), Minaei-Bidgoli *et al.* (2003), Watanabe and Nakayama (2003), Zhou (2003), Zang and Lin (2003), Kokol *et al.* (1999), and many more. Various data mining techniques have been applied in solving their problems.

For example, Ha *et al.* (2000) showed the use of web usage mining to enhance distance education. Their result showed that the effectiveness and efficiency of distance learning could be improved along two dimensions. Discovery of aggregate and individual paths for learners engaged in distance education could help in the development of effective customized education, and give an indication of how to best organize the educator organization's Web space. The discovery of association rules could make it possible for Web-based distance educators to identify knowledge structures and to reorganize web based on these structures. Lo and Ng (1999) studied the use of data mining to discover the access patterns of WWW users. They studied two approaches of association mining to uncover the pattern. These approaches are Extended Apriori algorithm (EAA) and OrderPattern Mining (OPM). Results showed that OPM was better in terms of time.

Students Enrollment

A vast amount of research indicates that factors influence on students enrollment such as, Bell and Fritz (1992) signified that deterrents to female enrollment in secondary agricultural education programs in Nebraska resulted from require of career information, career opportunities and nontraditional employment. In other hand, Bell and Fritz (1992) identified and defined deterrents to male enrollment in secondary agricultural education programs. They compared the results with females in a similar study. The result showed that counseling influenced by the perception of gender appropriate occupational roles and is a phenomenon of acceptance. Additionally, Shelton (2001) studied also reasons parents' state as the initial factors leading them to place their children in Protestant Christian schools (PCs), factors that lead parents to continue to enroll their children in the PCs. Parents and principals were asked to rank what they perceived to be the top five factors leading to enrollment of students in PCs. Ranking the various factors isolated the most common factors determining parents' decisions. Isolating the most-often chosen of the various factors aids the Protestant Christian, other nonpublic, and public schools better to evaluate how each attracts, maintains, and services their respective clientele.

Esters and Bowen (2004) also identified factors influence students to enroll in an urban agricultural education program. They concluded that parents or guardians were the primary individuals influencing their decision to enroll in an urban agricultural education program, although, the mother or female guardian were the most influential individual. Meanwhile, Stage and Hossler (1989) reported that father and mother educational achievement, as well as family income, are important factors affecting parents' educational expectations for their children.

Spohn, *et al*., (1992) revealed that level of parental educational attainment was a factor that influenced whether students could navigate the college application process and whether they witnessed firsthand the benefits of higher education. Similarly, siblings' college attendance influenced enrollment, because older siblings are often role models for their younger brothers and sisters. Low family income and the family's inability to help finance higher education emerged as inhibiting factors. High school personnel in Appalachia perceived a lack of parental encouragement for students to attend college.

Therefore, an example for course enrollment, Hall (2006) studied enrollment students in Practicum course. The purpose of the course is for the student to develop individual counseling skills while functioning in a counseling setting. Additionally, individual skills, students are also encouraged to participate in group counseling. The result showed that students enroll in practicum for the level most closely related to their professional goals.

Jm (2002) provided some insight on ways to attract new people to laboratory science and as a source of information for institutions to encourage young people to enter the field of laboratory medicine. Recruiters should focus marketing the program in their local geographical area. Family and friends should continue to spread the word about the profession. The ability to grow into other professions should be emphasized. College advisors should always remain visible and market the programs; and high school students should be exposed to the profession.

Among other factors discussed, peers also influence academic achievement in positive and negative ways, and for many students of lower socioeconomic status, academic success may be viewed contemptuously (Phelan *et al*., 1991).

Print ISSN 2306-9309 Online ISSN: 2306-9295

Table 2.4: studies on students' enrollment.

STUDENTS' ENROLLMENT		
Author	**Year**	**Title**
Hall	2006	*Psychology & Research in Education.*
Esters and Bowen	2004	Factors Influencing Enrollment In An Urban Agricultural Education Program
Jm	2002	Factors influencing student enrollment in clinical laboratory science programs
Shelton	2001	Factors Influencing Enrollment In Virginia's Protestant Christian Schools
Bell and Fritz	1992	Deterrents to female enrollment in secondary agriculture education programs in Nebraska
	1992	A Comparison of deterrents to nontraditional male and female enrollment in secondary agricultural education programs in Nebraska
Spohn *et al.*,	1992	Appalachian access and success: A research project of the Ohio Board of Regents and a consortium of two- and four-year colleges and universities in Appalachian Ohio
Phelan *et al.*,	1991	Students' multiple worlds: Negotiating the boundaries of family, peer, and school cultures
Stage and Hossler	1989	Differences in family influences on college attendance plans for male and female ninth graders

Apriori Algorithm

The approach based on Apriori algorithm has become one of the most common data mining methods, which was proposed by Agrawal *et al.*, (1994). While, Xu *et al.*, (2005) defined that Apriori algorithm is one of the earliest for finding association rules. This algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules. Additionally, (Dunham, 2003) identified that the basic idea of the Aprori algorithm is to generate candidate itemsets of a particular size and then scan the database to count these to see if they are large.

There are many applications that apply Apriori algorithm to get the related rules between data in large databases; one of the applications proposed is a new algorithm named Multipass with Inverted Hashing and Pruning (MIHP) for mining association rules between words in text databases. Two well known mining algorithms, the Apriori algorithm and the Direct Hashing and Pruning (DHP) algorithm, were evaluated in the context of mining test databases, and were compared with the

Print ISSN 2306-9309 Online ISSN: 2306-9295

proposed MIHP algorithm. Results showed that Apriori rules gave better results for small text databases (Holt and Chung, 2002).

In another approach, Chan *et al*. (2003) developed a new pruning strategy based on utilities that allow pruning of low utility itemsets to be done by means of a weaker but anti-monotonic condition. Their experimental results showed that their algorithm did not require a user specified minimum utility and hence was effective in practice. However, they showed Objective Oriented Association mining with top-k utility frequent closed patterns was feasible with extensions to the Apriori algorithm. Another work presented by Fernandez *et al*. (2001) also showed that Apriori algorithm produced interesting results. In their research, Apriori algorithm was applied to reduce the number of condition attributes and the number of decision rules obtained through rough sets. Their results highlighted the fact that in all cases the approach enhanced the output of the classification algorithm. Additionally, Cunningham and Frank, (2000), presented work which is to discover library reuse patterns in user-selected applications. Specifically, it considered the problem of discovering association rules that identify library components that are often reused in combination by application components.

Yan *et al.* (2005) presented a novel fuzzy comprehensive evaluation approach Based on Apriori algorithm for unit's bidding ability assessment in electricity market. The weight's value was determined by expert's knowledge in normal fuzzy comprehensive evaluation model, In proposed model, the weight's value of multi-factors was adjusted by improved priori algorithm to find association among many factors. Then, this method applied to evaluate unit's bidding ability of electricity market. A set of judgment factors includes SMP, market demand, the bidding price, minimal and maximal output of the units, characteristics of costs, time constraint on starting up and cease in this model. The numerical results showed that the proposed approach is right and effective. This method can precisely demonstrate the bidding units' technological and economical characteristic.

Pirttikangas *et al.* (2004) reported and analyzed the connection of data mining algorithms and routine learning in a mobile environment. Data collected from various ubiquitous sensors are used in recognizing and defining contexts, association rules determined the routines. The focus was on testing the suitability of the Apriori algorithm for this application area. Several useful routines were derived from the user data, and the results showed it is possible to utilize data mining in a mobile environment. They have verified that it is possible to associate simple contexts or actions to these locations using the (simplified) methodology of the Apriori algorithm. They built a prototype for an architecture, which aided in building interactive mobile services that adapt to the user's situation.

Table 2.6: Applications of Apriori algorithm.

APRIORI ALGORITHM		
Author	**Year**	**Title**
Toshev *et al*	2006	An Apriori-based Method for Frequent Composite Event Discovery in Videos
Xu *et al.*	2005	Mining Association Rules with New Measure Criteria
Yan *et al.*	2005	A Novel Fuzzy Comprehensive Evaluation Approach Based Apriori Algorithm for Unit's Bidding Ability Assessment
Chen *et al.*	2005	Utilize Fuzzy Data Mining to Find the Travel Pattern of Browsers
Varde *et al.*	2004	Apriori Algorithm and Game-of-Life for Predictive Analysis in Materials Science
Pirttikangas *et al.*	2004	Routine Learning: Analyzing Your Whereabouts. Information Technology
Chan	2003	Mining High Utility Itemsets
Lu *et al.*	2003	A Prediction Method of Fuzzy Association Rules
Dunham	2003	*Data Mining Introductory and Advanced Topics*
Holt and Chung	2002	Mining Association Rules in Text Databases Using Multipass with Inverted Hashing and Pruning
Sharagai & Schneider	2001	Discovering quantitative associations in databases
Fernandez *et al.*	2001	Minimal Decision Rules Based On The Apriori Algorithm
Cunningham and Frank	2000	Market Basket Analysis of Library Circulation Data
Brin *et al.*	1997	Beyond Market Basket: Generalizing Association Rules To Correlations
Agrawal *et al.*	1994	Fast algorithms for mining association rules in large databases

Methodology

In this study, the Knowledge Discovery in Databases (KDD) process methodology adopted from (Kuonen, 2003) is implemented. This methodology has been applied by several researchers which include Dunham, (2003), Düntsch *et al.* (2000), Fayyad

Print ISSN 2306-9309 Online ISSN: 2306-9295

(1996), Williams and Huang (1996), and Yacoben and Carmichael (1997). The phases and processes are illustrated in Figure 3.1.

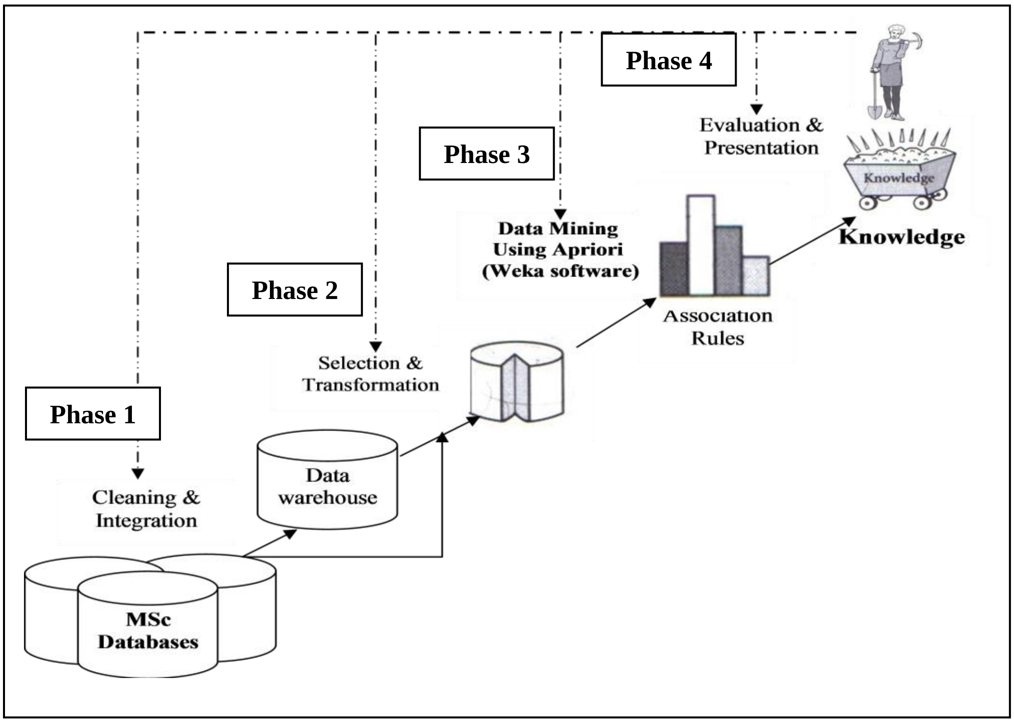

Figure 3.1: The KDD process and its phases.

In this study, the sampling obtained from distributed questionnaire comprises of 207 postgraduate's records who currently undertaking program that include MBA(Acct), MBA(General), MBA(Tourism), MHRM, MSc(Finance), MSc(IT), MSc(IS), MSc(ICT), MSc(Mgt), MSc(Edu Mgmt), MSc(Acct), MSc(Banking), MSc(Int Acct), MSc(Mgmt) and MSc(Mgt).

The dataset consists of demographic, socio-economic, and interest background of the students. The dataset is then processed and transformed through following phases:

Data cleaning

Basic operations such as the removal of noise or outliers if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, accounting for time sequence in-formation and known changes. There are some missing values found in the dataset, for the information about institute provides laboratory column, difficulty communicating with Malaysian

Print ISSN 2306-9309 Online ISSN: 2306-9295

column, and suitability accommodation column. For this study, the missing values were replaced with null.

Yi and Zhang (1998) said that, the way it expresses missing numeric values and character values are totally different. A missing numeric value is usually expressed as a period (**.**), but it also can be stated as one of other 27 special missing value expressions based on the underscore (_) and letters A, B,…,Z, that is, **._**, **.**A, **.**B,…, **.**Z.

For example, a missing value can be ignored, which is the easiest way to handle missing values, but it does not contribute to enhance the quality of the database that has the missing values. Another typical way to handle missing values is to replace them by using the average value, substituting with the maximum value, or using the minimum value. Other way, use a global constant to fill in the missing value: e.g., “unknown” and use the most probable value to fill in the missing value, (Kodratoff, 2000).

Data selection

This is the second step, which is selecting a data set, or focusing on a subset of variables or data samples, on which discovery is to be performed. Firstly, all the attributes has trained by Apriori algorithm and then we deleted the program of study which is Master and the type of the program which is coursework because this study for master students who are taking coursework program.

Data transformation

Where data are transformed or consolidated into appropriate forms, it can involve the following: smoothing which works to remove the noise from data such techniques include binning, clustering and regression, and aggregation where summary or aggregation operations are applied to the data. For example, the daily sales may be aggregated to compute monthly and annual total amounts.

Generalization of the data, where low level are replaced by higher level concepts through the use of concept hierarchies. For example, Categorical attributes: age may be mapped to young, middle-aged, and senior. In this study, the values of variables age, salary prior to study and expected salary have been transformed to the followings:

Print ISSN 2306-9309 Online ISSN: 2306-9295

Table 3.1: shows original variables' values and transformed variables' values.

Variable	Original	Transformed	Original	Transformed
Age	20 – 25 years old	Ag	26 – 30 years old	Bg
	31 – 35 years old	Cg	36 – 40 years old	Dg
	Above 40 years old	Eg		
salary prior to studying	Less than RM1,000	Sp	RM2,000 - RM3,000	Sp2
	RM1,000 - RM2,000	Sp1	Above RM3,000	Sp3
expected salary upon completion	RM1,000 - RM2,000	Su	Above RM3,000	Su2
	RM2,000 - RM3,000	Su1		

Data mining

In this step, essential processes where intelligent methods are applied in order to extract data association rules**.** In other word**,** deciding whether the goal of the KDD process is classification, regression, clustering, etc. In this study, the mining process is done using Apriori algorithm. Apriori algorithm is one of the most common used algorithms in association rules. Before discussed about the Apriori algorithm, let have an idea about the association rules mining. Lee and Suh, (1998) said that association rules are very useful for market database. The discovery of association rules is composed of two steps. The first step is to discover all of the frequent itemsets and the second step is to generate the association rules that have confidence values greater than the minimum confidence threshold from these frequent itemsets. An association rule consists of a condition and a result, and is usually represented in the form of *"if condition then result"*. The association rules are always defined on binary attributes and have an implication of the form X => Y meaning that data tuples for which the attributes of X have true values; the attributes of Y also tend to have true values. For example; suppose that people who purchase bread and cheese also purchase milk. In this case, the corresponding association rules is *"bread ^ cheese => milk"*. The antecedent of the rule X consists of bread and cheese and the consequent Y consists of milk alone. Say that a rule X => Y holds in the transaction set D with confidence *c%* if *c%* of transactions in D that contain X also contains Y. The rule X => Y has support *s%* in the transaction set D if *s%* of transactions in D contains $X \cup Y$.

Print ISSN 2306-9309 Online ISSN: 2306-9295

Association rules evaluation

Evaluation to identify the truly interesting association rules representing knowledge based on some interestingness measures. Also, it's selecting method(s) to be used for searching for patterns in the data. The same dataset was then tested using statistic to see the correlation between attributes.

Knowledge presentation

This is the last step, where visualization and knowledge representation techniques are used to present the mined knowledge in this study. Also, incorporating this knowledge into the performance system, or simply documenting it and reporting it to interested parties. This also includes checking for and resolving potential conflicts with previously believed (or extracted) knowledge. This data presented as documenting and reporting.

Results for Apriori Algorithm

Apriori is an algorithm that commonly used in association rule mining. An association rules (Chen, 2001) is intended to capture to certain type of dependence among items represented in the database. The standard measures to assess association rules are the *support* and the *confidence* of a rule, both of which are computed from the *support* of certain item sets. As we know, the rule of A=>B explains the confidence. Support, *s%*, is A and B occur together in at least *s%* of the basket and confidence, means that of all the basket containing A, at least *c%* also contain B. Minimum support and minimum confidence is needed to remove the unimportant association rules. The association rule is hold when it is greater than the minimum support and minimum confidence. Like other basket data problems, while apparently similar, have requirements that the support – confidence framework does not address (Brin *et al.,* 1997). The support factor, s, is generally specified by the user, however, there are no criteria to define it (Lee and Suh, 1998). Table 4.1 lists the number of rule sets generated using Apriori algorithm in the same data. A column in the table shows the different minimum confidence values and the different minimum support values presented in a row.

Table 4.1: The number of Rule Set From Apriori Algorithm

Sup (%) / Conf (%)	10	20	30	40	**50**	60	70	80	90
50	106	73	70	57	52	43	17	5	3
60	133	138	118	113	118	43	17	5	3
70	196	185	188	198	124	40	14	2	2
80	219	210	281	269	111	36	11	1	1
90	126	102	143	95	19	3	0	0	0

Print ISSN 2306-9309 Online ISSN: 2306-9295

Most literature on data mining suggests that the suitable rule be chosen at highest confidence. From Table 4.1, the support values in last three columns equal to 0. The support factor 60% at the highest confidence value generated 3 rules whish their accuracy ranging from 90.6% to 90.9. This is the reason to express the number of rule sets with support factor 50% with confidence factor 90% which has 19 rules with accuracy ranging from 90.2% to 93.3%. The 19 rules have higher accuracy comparing with the rules with support factor 60%. From Table 4.1, the selected minimum support and minimum confidence is 50% and 90%, which have 19 rules. This support and confidence is selected because of the highest rules accuracy. The other range for support and confidence is not selected because of the number of the rule that they are produced.

Table 4.2: Association Rules Obtained

NO	Rules	Support (%)	Confidence (%)
1	Ao5 <- SF, FT, YesFeo	50.5	93.3
2	Ao5 <- FT, Lecturer	51.0	92.5
3	Ao5 <- SF, FT, M	53.8	92.0
4	Ao5 <- SF, FT, Islam, M	53.4	91.9
5	Ao5 <- S, On, FT	51.9	91.7
6	Ao5 <- S, FT	55.8	91.4
7	Ao5 <- S, On, FT, Islam	50.0	91.3
8	Ao5 <- SF, On, FT	54.8	91.2
9	Ao5 <- S, FT, Islam	53.8	91.1
10	Ao5 <- On, FT, YesFeo	53.8	91.1
11	Ao5 <- SF, On, FT, Islam	53.4	91.0
12	Ao5 <- SF, FT	63.5	90.9
13	Ao5 <- SF, FT, Islam	62.0	90.7
14	Ao5 <- FT, YesFeo, M	51.4	90.7
15	Ao5 <- On, FT, YesFeo, Islam	51.9	90.7
16	Ao5 <- FT, YesFeo	61.5	90.6
17	Ao5 <- FT, YesFeo, Islam, M	50.0	90.4
18	Ao5 <- SF, On	59.1	90.2
19	Ao5 <- FT, YesFeo, Islam	59.1	90.2

Based on Table 4.2, all rules that were produced illustrate a logical relationship between the items, considered to be suitable rules. The rules are ranked according to confidence values column from highest confidence value to the lowest.

The highest percentage of confidence is rule no.1 which is at 93.3%. This means that students who found out about education opportunities in UUM from a friend, they are interested to study full-time and their financial source are self financed is enrollment in UUM is at 50.5%. In rule no.2, if the students were lecturers and they came to study full-time, they also tend to enrollment in UUM at least 51.0% and the confidence for this rule is 92.55%. For rule no.3, male students who prefer to study full-time and their financial source is self financed, they are interested to enrollment in UUM at 53.8% and the confidence for this rule is 92.0%. In rule no.4, male Moslem students who prefer studying full-time and their supplying by self financed, completely, they are enrollment in UUM at 53.4% and confidence at 91.9%. Rule no.5, shows the support if students study full-time, live on campus and are single, they likely to enrollment in UUM at 51.9% and the confidence for this rule is 91.7%.

From rule no.6, students like to enrollment in UUM who are single, study full-time at confidence 91.4% and support is 55.8%. In rule no.7, male Moslem students who prefer studying full-time, live on campus and are singles, they likely to enrollment in UUM at confidence 50.0% and support is 91.3%. For rule no.8, students who live by self financed, reside on campus and study full time, they likely to enrollment in UUM at least 54.8% and the confidence for this rule is 91.2%.

The same percentage of confidence and support is in rule no.9, rule no.10 which is at 91.1%, 53.8% respectively. This means that in rule no.9, single students who are Moslem and likely to study full-time, they are probability like to enrollment in UUM. While in rule no.10, students who live on campus, study full-time and found out about education opportunities from their friends, they likely to enrollment in UUM. For rule no.11, , Moslem students who prefer studying full-time, live on campus and their supplying by self financed, completely, they are enrollment in UUM at 53.4% and confidence at 91.0%. The highest percentage of support is rule no.12 which is at 63.5%. This means that students who study full-time and their financial support is self financed, they like to enrollment in UUM at 90.9%.

The same percentage of confidence is in rule no.13, rule no.14 and rule no.15 which is at 90.7%. This means that in rule no.13, Moslem students who like to study full-time and their support is self financed, they are interested to enrollment at 62%. While in rule no.14, male students who prefer to study full-time and they found out about UUM from friends, they like to enrollment in UUM at 51.4%. Whereas for rule no.15, students who are Moslem, live on campus, got information about UUM from friends and prefer to study full-time, they like to enrollment in UUM at 51.9%.

For rule no.16, full-time students who got information about UUM from friends, they like to enrollment in UUM at least 61.5% and the confidence for this rule at 90.6%. In rule no.17, male Moslem students who study full-time and got information about education in UUM from friends, they like to enrollment in UUM at the lowest support value which is 50.0% and the confidence for this rule is 90.4%.

The same percentage of confidence and support is in rule no.18, rule no.19 which is at 90.2%, 59.1% respectively. This means that in rule no.18, students who live on campus and their support is self financed, they are interested to enrollment in UUM. While in rule no.19, moslem students who got information about education in UUM from friends and like to study full-time, they like to enrollment in UUM.

Table 4.3: Association Rules Obtained

Factors	Occupation	Gender	Marital status	Mode of study	Friend told me about UUM	Living	Religion	Financial
Rule1				✓	✓			✓
Rule2	✓			✓				
Rule3		✓		✓				✓
Rule4		✓		✓			✓	✓
Rule5			✓	✓		✓		
Rule6			✓	✓				
Rule7			✓	✓		✓	✓	
Rule8				✓		✓		✓
Rule9			✓	✓			✓	
Rule10				✓	✓	✓		
Rule11				✓		✓	✓	✓
Rule12				✓				✓
Rule13				✓			✓	✓
Rule14		✓		✓	✓			
Rule15				✓	✓	✓	✓	
Rule16				✓	✓			
Rule17		✓		✓	✓		✓	
Rule18					✓	✓		
Rule19				✓	✓		✓	

Print ISSN 2306-9309 Online ISSN: 2306-9295

There are several influencing factors, which can influence student to enrollment in UUM. From Table 4.2, shows eight patterns that are influencing students and mode of the study is the most influencing pattern, (financial, religion and information source about education in UUM) patterns.

Thus, Rules extracted that can be used in knowledge base are as follows:

If student (Self Financial, Full time, Got information from friends) then enrollment

If student (Full Time, Lecturer) then enrollment

If student (Self Financial, Full Time, Male) then enrollment

If student (Self Financial, Full Time, Moslem, Male) then enrollment

If student (Single, On campus, Full Time) then enrollment

If student (Single, Full Time) then enrollment

If student (Single, On campus, Full Time, Moslem) then enrollment

If student (Self Financial, On campus, Full Time) then enrollment

If student (Single, Full time, Islam) then enrollment

If student (On campus, Got information from friends) then enrollment

If student (Self Financial, On campus, Full Time, Moslem) then enrollment

If student (Self Financial, Full Time) then enrollment

If student (Self Financial, Full Time, Moslem) then enrollment

If student (Full time, Got information from friends, Male)then enrollment

If student (Full Time, Got information from friends) then enrollment

If student (Full Time, Got information from friends, Moslem, Male) then enrollment

If student (Self Financial, On campus) then enrollment

If student (Full Time, Got information from friends, Moslem) then enrollment

Results For Statistical Method

In this part, statistical method is used to make comparison based on the results that obtained from Apriori algorithm. This method has used in SPSS tool to the analysis. To run the same data using SPSS tool, the data has formatted to be suitable with the tool. Table 4.4 shows results of correlation for all patterns and their relationships

Print ISSN 2306-9309 Online ISSN: 2306-9295

Table 4.3: The Result of correlation Between Items

Pearson Correlation	Year	Gender	Marital Status	Living	Religion	Financial Source	Occupation Before Study	Mode Of study	Friends Educational Opportunity
Year	1.000	.043	-.052	-.272	.015	-.032	-.210	.022	-.032
Gender	.043	1.000	-.012	.172	.130	-.228	.037	.242	.109
Marital Status	-.052	-.012	1.000	-.244	-.051	.236	-.153	-.100	-.064
Living	-.272	.172	-.244	1.000	-.014	-.096	.181	.347	.105
Religion	.015	.130	-.051	-.014	1.000	-.080	-.083	.041	.034
Financial Source	-.032	-.228	.236	-.096	-.080	1.000	.033	-.051	.081
Occupation Before Study	-.210	.037	-.153	.181	-.083	.033	1.000	.184	-.073
Mode of Study	.022	.242	-.100	.347	.041	-.051	.184	1.000	.030
Friends Educational Opportunity	-.032	.109	-.064	.105	.034	.081	-.073	.030	1.000

Based on the Table 4.3, the results show that is relation between the most influencing patterns this is mode of study and the enrollment for students whose the correlation coefficient is 0.022. In addition, the relation between gender and the enrollment for students whose the correlation coefficient is 0.043. From the table, there is high negative relation between place of living and the enrollment of students which is -0.272. Moreover, there is high negative relation between occupation before study and the enrollment for students whose the correlation coefficient is -.210. The correlation coefficient for marital status and students' enrollment is -0.052.

The relationship between students' religion and the students' enrollment is very low whose value is 0.15. The correlation coefficient for Occupation before study, information about UUM's education from friends and the enrollment for students is same whose value is -0.032.

Discussions of The Results

Based on the experiments that have been undergone, there are a few factors that influence the students to select UUM as their choice for place of study. These factors can be illustrated in figure 4.1

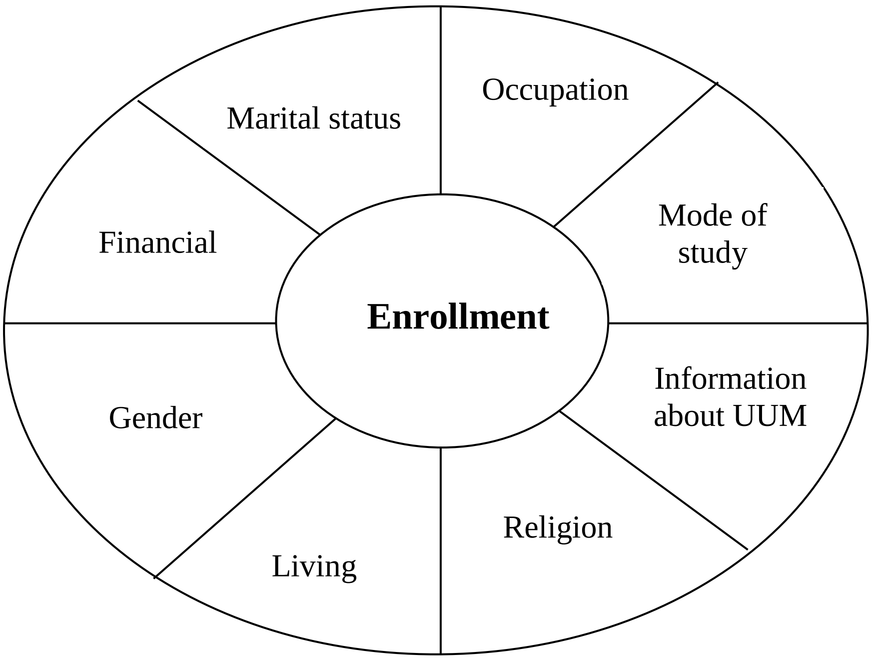

Figure 4.1: Influencing factors

From figure 4.1, it shows the relation between students' enrollment and influencing factors:

a. mode of study:
From the rules generated mode of study is the most influence factors that most of students' enrollment in UUM is full-time students.

b. Information about UUM's education from friends
This factor is influence students' enrollment as well where; most of students who registered in UUM have found out about education in UUM from friends.

c. Religion
In UUM's students are several religions, and this factor influence students to select UUM as well.

Print ISSN 2306-9309 Online ISSN: 2306-9295

d. Living

This factor is influence students' enrollment as well where; most of students who registered in UUM are living on campus.

e. Financial

From the experiments that have done, the majority of students who like to enrollment in UUM is supported by self financed.

f. Marital status

There is no high relation between this factor and students' enrollment, but from the experiments most full-time students who live on campus are single.

g. Gender

From the experiments, most of Moslem students who got information from friends and their financial source is self financed, they are male.

h. Occupation

This factor doesn't have high relation with students' enrollment, but extracted rules shows that students who had work before enrollment in UUM to study master are interested to study full-time.

Conclusion

Data mining is one of the branches of study in Artificial Intelligence discipline. Data mining techniques have been widely accepted to handle an abundance set of data. The capability of data mining to assist in analyzing data, beyond descriptive task, has intrigued researches to apply data mining techniques in their studies.

Like mining for gold, data mining sifts through large databases and extracts a wealth of customer data, which can then be turned into usable, predictive information (Goolsby, 2005).

Data mining is a term used to describe the process of analyzing an abundance set of data. The data is analyzed, organized and summarized into useful information. Data mining concept is intelligent enough to describe data and forecast the possibilities. It is also known as *knowledge discovery* due to the existence of this intelligent element in data mining. Data mining widely accepted in various domains such as business

Print ISSN 2306-9309 Online ISSN: 2306-9295

(Rygielski *et al.*, 2002), finance (Zhang and Zhou, 2004), engineering (Gardner and Bieker, 2000), students' enrollment (Esters and Bowen, 2004), education (Ma *et al.*, 2000) and etc. In this study, association rules have been applied to students' data.

It is important to mention that Association rule mining is very useful in order to find the factors that influence students to enrollment. This study also uses statistical method which is correlation coefficient between variables in order to present the relation between factors and not to compare with apriori algorithm. Results indicated that Apriori algorithm is able to identify influencing factors. The most obvious effect is to know influencing factors, which influence students to enrollment that can increase the number of students in UUM.

Apriori algorithm is most and common association rule algorithms, suitable in discovering relationship between data (Agrawal *et al*., 1994). The generated rules can easily be understood. To mine students' data, there is limitation, which is the kind of data needed to train. It is necessary to have a large number of real transactions to get meaningful data.

Based on the results those have been produced by Bogerlt apriori software using Apriori algorithm, a suggestion about software can be used. It could be better if train the same data with other software such as Weka software and also other association rule's algorithm such as Sampling algorithm. In addition, the study should be comparison between more than algorithm. Although data mining seems to be very intriguing concept, it should be used with carefully outlined criteria, which should not be based on any level of prejudice or stereotype assumption. The factors can help UUM to make good promotion to students who are likely to be enrollment.

References

Agrawal. R, Srikant. R. (1994). Fast algorithms for mining association rules in large databases. In: Proc. of the 20th Intel. Conf. on Very Large Data Bases, Santiago, Chile, pp. 487-499, Sep.1994.

Antonie, M-L., Zaiane, O. R., & Coman, A. (2001). Application of Data Mining Technologies for Medical Image Classification. *Proceedings of The Second International Workshop on Multimedia Data (MDM/KDD'2001).*

Bansal, K., Vadhavkar, S. and Gupta, A. (2001). *Neural Network Based Data Mining Applications for Medical Inventory Problems.* [Online]Available.*http://scannergroup.mit.edu/Project%20graphics/dmpaper1.pdf.*

Bariani, M., Cucchiara, R., Mello, P. and Piccardi, M. (1999). Data Mining For Automated Visual Inspection.

Bell, L. C. & Fritz, S. M. (1992). Deterrents to female enrollment in secondary agriculture education programs in Nebraska. *1992*, *34*(4), P.P.39-46.

Bell, L. C. & Fritz, S. (1992) A Comparison of deterrents to nontraditional male and female enrollment in secondary agricultural education programs in Nebraska.1992, 35, No. 4.

Blazejewski, A., and Coggins, R. (2004). Application of Self-Organizing Maps to Clustering of High-Frequency Financial Data. *The Australasian Workshop on Data Mining and Web Intelligence (AWDM&WI2004)*

Brin, S., Motwani, R., & Silverstein, C.,(1997).Beyond Market Basket: Generalizing Association Rules To Correlations.ACM SIGMOID-'97.

Chan, K. P., Fan, w., Prodromidis, A, L., and Stolfo, S, J. (1999). Distributed Data Mining in Credit Card Fraud Detection. *Florida Institute of Technology*, *P.P.* 67-74.

Chan, R., Yang, Q., and Shen, Y. (2003). Mining High Utility Itemsets. *Data Mining, ICDM 2003. Third IEEE International*, pp. 19 - 26

Chen, M.-S., Jan, J., Yu, P.S. (1996). *Data Mining: An Overview from a Database Perspective*. IEEE Transactions on Knowledge and Data Engineering, (8:6). Pp 866.883

Chen, Q., Jianghong, H., He, W., Mao, K., and Yungang Lai. (2005) Utilize Fuzzy Data Mining to Find the Travel Pattern of Browsers. Computer and Information Technology, 2005. CIT 2005. The Fifth International Conference on. xxv+1148, P.P.228-232.

Chen, S. (2001). *Strategic Management of e-Business*. Sussex: John Wiley & Sons, Ltd..

Cunningham S. J., and Frank E. (2000) Market Basket Analysis of Library Circulation Data, in Proc. Of the Sixth International Conference on Neural Information Processing, Vol. 2, Perth, Australia, IEEE Service Canter, pp. 825-830.

Delavari N, Beikzadeh M. R. (2004). A New Analysis Model for Data Mining Processes in Higher Educational Systems, MMU International Symposium on Information and Communications Technologies 2004, Putrajaya, Malaysia, 7th-8th October 2004.

Doddi, s.,Marathe, A., Ravi,S.S. and Toney,D.C.,(2002). Discovery of Association Rules in Medical Data. Online: http://www.c3.lanl.gov

Dunham, M.H. (2003). Data Mining Introductory and Advanced Topics. New Jersey: Prentice Hall.

Fernandez, M, C., Menasalvas, E., Marban, O., Pena, J, M., and Millan, S. (2001). Minimal Decision Rules Based On The Apriori Algorithm. *IEEE*, 3, pp. 691-704.

Gardner, M. and Bieker, J. (2000). Data Mining Solves Tough Semiconductor Manufacturing Problems. In *Proceedings Of The Sixth ACM SIGKDD International Conference On Knowledge Discovery And Data Mining.*

Goolsby, K., (2005). About-Business.org. http://about-business.org/index.php

Grossman, R., Kasif, S., Moore, R., Rocke, D., & Ullman, J. (1998) A Report of three NSF Workshops on Mining Large, Massive, and Distributed Data, http://www.ncdm.uic.edu/m3d2.htm.

Ha, S. H., Bae, S. M., and Park, S. C. (2000). Web mining for distance education. *Proceedings of IEEE International Conference management of innovation and technology, 2000,* 2, pp. 715-729.

Hall, Joseph R, P. (2006). *Psychology & Research in Education.*

http://www.soe.ku.edu/pre/courses.php

Holt, J. D., Chung, S. M. (2002). Mining Association Rules in Text Databases Using Multipass with Inverted Hashing and Pruning. *Proceedings of the 14th IEEE International, 2002*, pp. 49 – 56.

Jm, S. (2002) Factors influencing student enrollment in clinical laboratory science programs. 2002 Aug-Sep;21(7):31-3.

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12380029&dopt=Abstract

Kauderer *at al.* was interested on optimizing the collection efforts in automobile financing, while Tay and Shen applied data mining technique for financial investment.

Kauderer, H., Nakhaeizadeh, G., Artiles, F., Jeromin, H. (1999). Optimization of collection efforts in automobile financing a KDD supported environment. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining.*

Kock, A., Michaellis, S., Seger, J., & Miloucheva, I. Inter-domain traffic engineering using advanced visual data mining technology, http://www.ist-intermon.org.

Kodratoff, Y. (2000). About Knowledge Discovery in Texts: A Definition and an Example.

Kohavi, R., and Provost, F. (2001). Applications of Data Mining to Electronic Commerce. Data Mining and Knowledge Discovery, vol. 1, 1-7.

Kokol, P., Zazula, D., Brumec, V., & Kolenc, L. (1999) Nursing informatics education for the next millennium. Future Generation Computer Systems 15, pp. 211–216.

Lee, K. B., and Suh, S. C. (1998). The Efficient Algorithm in The Volume of Market Basket Data for Association Rules, *Expersys-98, The 10th International Conference*.

Lo, C., and Ng, V. (1999). Discovering web access orders with association rules. *Proceedings IEEE International, 6.*

Lu, J., Xu, B., and Jiang, J. (2003) A Prediction Method of Fuzzy Association Rules. Information Reuse and Integration, 2003. IRI 2003. IEEE International Conference on. Xl+620, P.P. 98 – 103.

Luan, J. (2002). Data Mining and Knowledge Management in Higher Education: Potential Applications. Presentation at *AIR Forum, Toronto, Canada.*

Ma, Y., Liu, B., Wong, C. K., Yu, P. S., Lee, S. M. (2000). Targeting The Right Students Using Data Mining. In *Proceedings Of The Sixth ACM SIGKDD International Conference On Knowledge Discovery And Data Mining.*

Ma, Y., Liu, B., Wong, C. K., Yu, P. S., Lee, S. M. (2000). Targeting The Right Students Using Data Mining. In *Proceedings Of The Sixth ACM SIGKDD International Conference On Knowledge Discovery And Data Mining.*

Min, H., Min, H. and Emam, A. (2002). A Data Mining Approach to Developing the Profiles of Hotel Customer. *Department of Computer Science & Engineering*, University of Loisville, USA.

Minaei-Bidgoli, B., A. Kashy, D., Kortemeyer, G., & F.Punch, W. (2003). Predicting Student Performance: An Application of Data Mining Methods with the Educational Web-Based System LON-CAPA. *Proceedings of 33rd ASEE/IEEE Frontiers in Education Conference, November, 5-8*.

Mitkas, P.A., Symeonidis, A. L., Kehagias, D., & Athanasiadis, I. (2003). Application of Data Mining and Intelligent Agent Technologies to Concurrent Engineering. *Proceedings of 10th International Conference on Concurrent Engineering (CE-2003), Madeira, Portugal.*

Print ISSN 2306-9309 Online ISSN: 2306-9295

Moser, S. A., Jones, W. T., and Brossette, S. E. (199). Application of Data Mining to Intensive Care Unit Microbiologic Data. The University of Alabama at Birmingham, Birmingham, Alabama, USA

Phelan, P., Davidson, A., & Cao, H. (1991). Students' multiple worlds: Negotiating the boundaries of family, peer, and school cultures. 1991, 22, P.P. 224-250.

Pirttikangas, S., Riekki, J., and Röning, J. (2004) Routine Learning: Analyzing Your Whereabouts. Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference, 2004. 2, P.P. 208 – 213.

Preston, J. A., and Shackelford, R. (1987). A System for Improving Distance and Large-Scale Classes. In *ITiCSE '98*, Dublin, Ireland.

Rygielski, C., Wang, J. C., & Yen, D. C. (2002*) Data mining techniques for customer relationship management.* Technology in Society, 2002. Vol. 24, P.P. 483-502.

Salazar, A., Gosalbez, J., Bosch, I., Miralles, R., and Verga, L. (2004). A Case Study of Knowledge Discovery on Academic Achievement, Student Desertion and Student Retention.

Scott, R.I., Svinterikou, S., Tjortjis, C. and Keane, J.A. (1996). Experiences of Using Data Mining in a Banking Application. *Department of Computers,* UK, 1-4.

Shelton, J. R. (2001) FACTORS INFLUENCING ENROLLMENT IN VIRGINIA'S PROTESTANT CHRISTIAN SCHOOLS. Blacksburg, Virginia, 2001.

Shragai, A., Schneider, M. (2001) Discovering quantitative associations in databases. IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th. Vol.1, P.P. 423 – 428.

Shragai, A., Schneider, M. (2001) Discovering quantitative associations in databases. IFSA World Congress and 20th NAFIPS International Conference, 2001. vol. 1, P.P. 423 – 428.

Smith, K. A., and Gupta, J. N. D. (2000). Neural Networks in Business: Techniques and Applications for the Operations Researcher. *Computer & Operations Research,* 27, P.P. 1023-1044.

Spohn, K., Crowther, T., & Lykins, C. D. (1992). *Appalachian access and success: A research project of the Ohio Board of Regents and a consortium of two- and four-year colleges and universities in Appalachian Ohio*. 1992.

Stage, F. K., & Hossler, D. (1989). Differences in family influences on college attendance plans for male and female ninth graders. *Educational Leadership and Policy Studies, Indiana University, 47405 Bloomington,* IN*, 30,* P.P.301-315.

Thiesing, F. M., Middelberg, U., and Vornberger, O. (1995) Short Term Prediction of Sales in Supermarkets. Proceedings ICNN'95, IEEE, Vol. 2, P.P. 1028-1031.

Tigrani, Vida and John, George H. (1998) Data Mining and Statistics in Medicine: An Application in Prostate Cancer Detection. *In JSM98, the Proceedings of the Joint Statistical Meetings, Section on Physical and Engineering Sciences.* American Statistical Association.

Toshev, A., Bremond, F., and Thonnat, M. (2006) An APRIORI-based Method for Frequent Composite Event Discovery in Videos. Computer Vision Systems, 2006 ICVS '06. IEEE International Conference. 2006, P.P. 10 - 10

Varde, A, S., Takahashi, M., Rundensteiner, E, A., Ward, M, O., Maniruzzaman, M., and Sisson Jr, R, D. (2004) Apriori Algorithm and Game-of-Life for Predictive Analysis in Materials Science.

Walker, P. R., Smith. B., Qing, Y. L., Famili, A. F., J. Valdes, J., Liu, Z., & Lach, B. (2004) Data Mining of Gene Expression Changes in Alzheimer Brain. *Artificial Intelligence in Medicine. 2004,* Vol. 31, P.P. 137-154.

Watanabe, T and Nakayama, M. (2003). Fuzzy Rule Extraction Based on the Mining Generalized Association Rules. Man and Cybernetics, 2003.IEEE, 3, P.P. 2690 – 2695.

Xu, Y., Zhou, Sen-Xin., and Gong, Jin-Hua. (2005). Mining Association Rules with New Measure Criteria. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005

Yan, H., Ma, R., and Tong, X. (2005). A Novel Fuzzy Comprehensive Evaluation Approach Based Apriori Algorithm for Unit's Bidding Ability Assessment. Natural Science Foundation of China

Yi and Zhang (1998) Handling Missing Values in the SQL Procedure**.** *Domain Solutions Corp., Cambridge, MA*

Zang, W., and Lin, F. (2003). Investigation of Web-based teaching and learning by boosting algorithms. *Proceedings of IEEE Information Technology: Research and Education, 2003,* P.P. 445-449.

Zhang, D., and Zhou, L. (2004). Discovering Golden Nuggets: Data Mining in Financial Application. *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews,* 2004. 34, P.P. 513-522.

Zhou, J. (2003). Discovering Association Rules In Engineering Document. *Natural Language Processing and Knowledge Engineering, 2003,* P.P. 339 – 344.

Zhou, Z. H., Jiang, Y. and Chen, S.F. (2000). A General Neural Framework for Classification Rule Mining. *Internal Journal of Computers, System, and Signals,* vol. 1, P.P. 154-168.