Mining of Egyptian Missions Data for Shaping New Paradigms

Laila Mohamed ElFangary
Information Systems Department
Faculty of Computers and Information - Helwan University
Cairo, Egypt
e -mail: lailaelfangary@gmail.com

Abstract—This paper reviews data mining applications of students' databases in educational institutions. Data mining techniques that predict and improve students ' retention rates and success is presented. Moreover, the Missions Administration at the Ministry of Higher Education in Egypt and previous analysis done on the missions databases is described. The paper further describes the methodology used for analyzing the database for the ministry of higher education in Egypt. The process starts by extracting a subset of data including the missioners and the mission's data, countries, specialties, departure and arrival dates and finally the extension requests from the missioners. These data were extracted into a data warehouse for the analysis purpose. The used model discovered the best and the worst countries for student mission. A detailed analysis discovered the best and the worst specialties in the previously discovered countries. Moreover, the analysis revealed the effect of the marital status on the mission of students in foreign countries. A visual display using a chart was used to express the information to business users. This model may help in achieving effective decision making in the reallocation of Egyptian students to other countries.

I. INTRODUCTION

The effective use of information technology in higher education requires good tactics as well as a sensible strategy. The reform of higher education through strategies based on database analysis will affect the overall performance of transitional countries and will shape new paradigms in higher education development. [1]

It is essential to evolve from current decision support environment in higher education systems to knowledge management driven data mining environment. [5] Studies have addressed the capabilities and strengths of data mining technology as applied to higher educational system. Several models were proposed using data mining techniques to enhance the quality of higher educational systems. [12] Data mining is the key aspect for extracting quality information to better understand college students' needs or behaviors and is critical to maintain successful education. [6] Knowledge that is hidden among the educational data set can be extractable through data mining leading to useful information of patterns, associations and significant structures. Data mining technology enhances decision making procedures on the various processes of higher education such as planning, counseling and evaluation. Data mining of higher educational systems will help maximize its efficiency by reducing the cost of system processes as it provides guidelines to increase students' promotion rate, success and learning outcome. [9]

II. DATA MINING OF STUDENTS DATABASES IN EDUCATIONAL INSTITUTIONS

There are several data mining techniques that could be applied to analyze the various students and student activities. Those techniques include classification to categorize students into different groups, prediction to predict students success, segmentation or cluster analysis to group similar students, association to identify courses that are taken together, sequence analysis or generators to find patterns and trends overtime. A data mining model was developed to explain IPEDS (Integrated Postsecondary Education Data System) graduation Rates. Analysis was done at institution level to mine both 2-year and 4-year graduation rates. A tree-based classification and prediction method with binary splits was applied to predict graduation rate. [4] Another model of seven higher educational systems processes that included evaluation, planning, registration, consulting, marketing, performance and examination was proposed to improve the efficiency and effectiveness of those processes. The model identifies the parts of the process that can be improved by data mining technology. [9]

Another case study was related to applying a predictive model in future admissions data sets for recruitment plan, conversion efforts, and other administrative decisions. A predictive model was developed that helps a school identify those candidates who are most likely to enroll and succeed. The Input variables were demographic information, academic scores and geographic and demographic features of home residence. Decision estimates were based on prior knowledge of success rates and target variable were derived from term GPA (If term GPA >=2.0). Fall 2001 data were used for model development and that of fall 2002 were used for validation scoring. Other analysis was used to apply the development of mathematically derived clusters learner typologies, using data mining based clustering technique. In 2002, Students demographics and enrollment records at Cabrillo College were clustered by race, gender and GPA the typologies developed provided insights into student learning and gave early warnings of trouble and were used for benchmarking and ongoing monitoring. [6]

Data mining techniques were applied to classify students in their predicted classes. The demographic information and course assessment records of 1360 students enrolled for five courses at the business studies department of Turks and Caicos Islands Community College from 2004 till 2006 were analyzed using Clementine 10.0 software applying the C5.0 algorithm to develop decision trees. The attributes that best represent the division of the training sets were determined and a neural network model was processed which reached an estimated accuracy of 99.57% contained 76 input layer neurons, 4 hidden layer neurons and 5 output layer neurons. The attributes with the highest relative importance attribute was that of Credit Performance per cumulative grade point average (CGPA) with value of 0.311. The Generalized Rule Induction was applied to the attributes of six subjectively selected clusters. Furthermore demographics were applied to the clusters attributes to produce association rules that identified patterns in the dataset. Performance prediction indicators that predict students' final achievement upon graduation were derived from analysis of tests and examination scores. Effective monitoring and reporting tools of students' academic performance were derived from rules produced applying data mining. [2]

III. DATA MINING TECHNIQUES THAT PREDICT AND IMPROVE STUDENTS' RETENTION RATES

Several case studies were applied to mine higher education. An analysis included predicting end of 1st year performance in a 4 year private higher education institution, the analysis included undergraduate students specialized in science, math and engineering. The freshman-sophomore retention rate was 90% and the 4-Year graduate rate was 80%. The study considered the factors have significant impact on student's 1st Year "academic success". Classification and regression trees were developed on the cleaned variables of demographics information and pre-college performance indicators. The cumulative grade point average (GPA) by the end of 1st year was predicted. There was a correlation between retention and graduation rates and high school Rank Percentile and SAT Math. [4] Weka open source software classification algorithms J48, was used as an analysis tool on data collected from on line survey filled by senior undergraduate students at the Faculty of Economics and Business Administration in Babe-Bolyai University, Cluj-Napoca, Romania. The survey questions evaluated the motivation of senior students to go on with graduate studies. Data related to their demographics, financial, academic and career achievements and opinions on various aspect of the education process were collected. The analysis of the data differentiated and predicted students' choice in continuing their education with post university studies for a master or PhD degree through decision trees. [16]

Data mining applications in higher education institutions include marketing, alumni fund raising, survival analysis and persistence. Data mining projects were initiated in enrollment management in both four year and two-year

institutions. Persistence, that is students who returned to college year after year until they graduate, was chosen as an indicator of academic performance and enrollment management. The analysis helped identify those students who are less likely to return from semester to semester in order to implement interventions and marketing strategies that would enhance the persistence rate. Persistence modeling helped reduce students' dropouts which resulted in increased efficiency, higher graduation rates, and lower cost to society as a whole.[5] Factors affecting bad behavior of students were analyzed and vocational education students were classified using several attribute selection techniques in order to develop appropriate strategies that prevented them from being expelled. Naive Bayes classifier, Baysian Belief Network, C4.5 algorithm and RIPPER algorithm were used to compare hybrid and single classification methods. The k-fold Cross Validation technique was used to measure classification efficiency. The results revealed that generally hybrid classification using genetic search as attribute selection and a classifier had higher accuracy rate than simple classification. The hybrid classification technique using genetic search and correlation based feature selection evaluator with C4.5 algorithm revealed the highest accuracy rate at 82.52%, yet using F-measure evaluation revealed that C4.5 algorithm did not fit for all data types. The F-measure showed that a better precision value can be given by the hybrid classification technique using genetic search and wrapper subset with Baysian belief. [14]

IV. DATA MINING TECHNIQUES THAT PREDICT AND IMPROVE STUDENTS' SUCCESS

In order to improve the quality of a university services success and failure factors and ways that increase students' success chances were discovered by analyzing pedagogical results. A goal oriented process design was developed by interviewing all those involved in the educational such as studies mangers, teachers and students who specified their needs and goals. This information was used to design a multidimensional model and a data mining module. The model revealed that there is not enough information of partial examination results, students' absence ratio, and students' personal information which were needed for effective decision making. A star schema was applied to identify measures provided by success and absence ratio and dimensions that correspond to the analysis axis of each measure. To examine presupposed associations such as to check if an information system course is a prerequisite for a database course one should examine how confident is the rule that students who have high grades in information systems also have high grades in database. Apriori Algorithm was used to implement the decision tool which was developed with Delphi 7.0 using Oracle 9i database software. The decisional data were organized in data cubes for Online Analytical Processing. The tool was tested on the Computer Science Department database and the different analysis axes were explored with regard to their influence on success or failure of students. The analysis enabled

discovery of the tendency of a particular teacher to give bad or good marks, verification of the impact of the students' absences on their exam results and difficulties the students had in certain modules as opposed to others. [11]

An enhanced analysis model (DM_EDU) was presented to apply data mining in higher education. Part of the model, the student assessment sub-process was implemented to asses students taking computer programming course at the Faculty of Engineering of Multimedia University (MMU) in Malaysia. This implementation revealed the importance of data mining in discovering useful patterns that improve the decision making process. As for example the faculty could take the necessary preventive actions based on patterns and predication of which students will relatively perform badly in this course. [10] The relationship between students' university entrance examination results and their success was examined by applying cluster analysis and k-means algorithm techniques. As university students were grouped according to their characteristics, forming clusters. The fields used in the study were transformed for data mining; they included the percentile the student fell into in the university entrance exam, and the grade they obtained, their gender, the type of high school and faculty. Successful partitioning using the K means algorithm was achieved with 5 clusters. The results showed that the most successful cluster with regard to grades is cluster 1 while the least successful was cluster 5. Most of the Students in cluster 1 were from the Faculty of Arts and Sciences who study hard to keep their scholarships thus have high success grades. The majority of the students in Cluster 5 were from the Faculty of Communication and Faculty of Business Sciences who had lower grades and results in the university entrance exam. The analysis provided varied and significant findings, which may lead to higher quality of education. [15]

Classification learning and data clustering techniques were used for analysis and prediction of students' performance on exams at the Faculty of Economics and Business Administration Cluj-Napoca. Data was extracted from students' filled online surveys and students' scholastic situation database that included attributes related to their gender, previous degrees, scholastic situation, types of gained scholarships, and interruption of study, exams absence, tuition, and students' opinions. A decisional tree in textual and graphic based on the scholastic attribute (exams success/failure) was presented. [17] Student Information System databases were analyzed using rough set theory and an algorithm was discussed to predict students who show weak performance in their academic program, their likelihood of success and graduation, and their needs of additional training. The technique discussed enabled accessing the performances of a particular student by accessing a group of students which can be effectively used in cases where certain attribute values of the students cannot be drawn or presented. The data sets included educational resources related to the student, information about their users

and activity log databases. Using rough set logic, series of pattern classifiers with various parameters were designed, implemented, and evaluated to compare students' performance. The student data information table consisted of three calculated attributes with sub attributes. This included student academic based on student performances in theory, practical, attendance, participating seminar, paper presentation, raising doubts, reading books, and participating department activities; student non-academics based on student performances in sports, various university clubs and activities; and human behavior relationships attributes based on performances in Faculty, class mates, and hobbies. A matrix of a set of instances represented the values associated to an attribute that is either numerical or nominal and decision attribute containing information about the decision to take upon the individual's consideration [12].

V.   MISSIONS ADMINISTRATION AT THE MINISTRY OF HIGHER EDUCATION IN EGYPT

The Missions Administration supports the missioners within the framework of the Ministry of High Education program for student missions in various specializations and including the following types: External missions for obtaining the doctor of philosophy degree, for a maximum period of four years (This category represents about 74% of the missions' total budget; Joint- supervision missions; Internal missions; Scientific missions for clinical medicine students who obtained the Ph.D. degree for a period of one year renewable for another year; Scientific missions financed by the Missions Administration for a time period of 3, 6 or 9 months; In addition to grants offered to the State, according to the executive programs. The fifth five- year plan of the Missions Administration covering the period 2002/2007 had a total budget of 2.5 billion EGP. The total number of missioners on study or scientific missions is 4720. As for the sixth five- year plan (2007/2012) of the Missions Administration the estimated budget is 2.8 billion EGP. (Including the cost of the missioners that are reserve or disqualified who are estimated to be 600 external missioners).

VI.   ANALYSIS OF THE MISSIONS DATABASES

Graduate study abroad comprise many challenges which include the absence of the learners from their homeland, non-conformity of foreign education with domestic needs, minds' emigration, and the cost of forming scientific cadres in foreign countries. Statistical analysis of the missioners' data revealed that minds' emigration is a major challenge facing the missions administration. Abstaining from returning home started to emerge as a clearer phenomenon in year 1963. After the June 1967 war, the number of missioners who had achieved their specialized studies and obtained the academic degrees they had been sent for and did not return home increased remarkably, especially in the United State of America. According to the study submitted

to the National Council for Education, Scientific Research and Technology in its twenty- first session in year 1993/1994, concerning the political aspects of scientific missions, the number of missioners to foreign countries who had abstained from going back to their homeland reached 1530 individuals during the period 1963 to July 1985.

Moreover a study revealed that the ratios of missioners abstaining from coming back home increases when the State's share in financing their mission decreases or does not exist. Hence the ratio of non- returning missioners was 32.9% in the case of personal grants; 21.4% in the case of foreign finance; and 21.1% in the case of the State's grants. Specialization also influences the ratio of non- returning missioners; with engineering sciences coming on top of the list with the ratio of 51%, followed by basic sciences (14.8%); medical and pharmaceutical sciences (12.5%); then all the other specializations. From the point of view of the hosting country, the U.S.A. registered the highest ratio of non-returning missioners (37.8%) followed by Canada (28.2%), the United Kingdom (21.2%), then France (6.8%), and Germany (6%).

During the five-year period (2003-2008), the number of non-returning missioners under all the financing categories reached 763 missioners in the U.S.A., whereas the number of those who completed their studies before the scheduled date was 4 missioners in the U.S.A. In the U.K., non-returning missioners were 186 individuals and those who completed their studies sooner than planned were 9 individuals. Hence, the number of non- returning missioners has been growing lately, entailing for the State the loss of the large sums of money spent on them, as well as the loss of the anticipated benefits from their gained knowledge. Under all kinds of missions, non-returning individuals numbered 1395 missioners, during the period (2003/8). The Ministry of Education has filed suits against them to claim the costs spent on them. The study showed that non-returning missioners at the end of their mission term are concentrated in the most advanced countries from the scientific and technological point of view, as those countries possess huge capabilities and facilities as well as material temptations that are unavailable in the missioners' homeland.

Most non-returning missioners (664 individuals) are among the beneficiaries of the State's public missions; while the least number of non-returning missioners (35 individuals) are among those benefiting from joint- supervision missions. Furthermore, the greatest number of non-returning missioners is concentrated in the U.S.A. (763 individuals), followed by 186 non-returning missioners in the U.K., 139 in Germany and 87 in France. On the other hand, the total number of non-returning missioners from the other countries is 130 individuals or only 17.3% of the number of non-returning missioners from the U.S.A. alone. The study concluded that the ratio of non-returning missioners increases among those delegated in study leaves as compared to the others; it also increases as the State's financing role decreases or disappears. Non-returning

missioners who have achieved their mission objective are mainly concentrated in the U.S.A. and Canada followed by the U.K., with the engineering sciences occupying first position on the list of non-returning missioners' specializations. The phenomenon of non-returning missioners after achieving the goal of their mission requires a careful analysis of the reasons behind such behavior and how it can be overcome. [3]

VII.   THE IMPLEMENTATION METHODOLOGY

1.   THE NATURE OF THE DATA

The data was extracted from the mission database. A subset of the data was extracted into flat files. These flat files were later imported into a data warehouse for use in the analysis process. This subset was classified into delayed and succeeded missioners. The delayed missioners are those who either exceeded five years abroad on their personal expense or they could not complete the mission due to any reason. They are considered as delayed missioners because both of them did not return back soon to serve their colleges/country for seven years which means that they postponed the strategy planned by the ministry of higher education to advance the educational process for the colleges in Egypt. The succeeded missioners are those who finished their mission in five years or less and returned back to Egypt. The data revealed that the number of missioners that were analyzed during a ten year span was 1661, of which 1303 are succeeded missioners while 358 are delayed. The data warehouse included the missioners marital status, the mission travel date, departure date, the extensions required for another year due to certain difficulties, the specialties, and the countries the missioners go to. Moreover other parameters were also included which are the Egyptian colleges and universities/institutions in which the missioners work. The number of universities/institutions is 26, while the number of colleges is 298. The objective is to study and analyze the delayed and succeeded missioners to help in implementing a missioners model.

IMPLEMENTATION METHODOLOGY

The analysis was done using the Cross Industry Standard Process for Data Mining (CRISP-DM) which is a data mining process model that describes commonly used approaches that expert data miners use to tackle problems [13]. The implementation was done using SQL (Structured Query Language).

According to CRISP-DM, a given data mining project has a life cycle of six phases [7], these phases are Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. The phase sequence is adaptive, that is the next phase in the sequence often depends on the outcomes associated with the previous phase [8]. The phases and the implementation are explained as follows:

1. Business Understanding:

This phase includes identifying the objectives and requirements of a case which is represented as a data mining problem definition in addition to an initial plan to achieve the identified objectives. Thus, the main business objectives and requirements for the business users were to evaluate the missioners process and to examine criteria that may affect the study of the missioners abroad, for instance the marital status. Also, the main goal was to identify and analyze the countries and specialties for the succeeded and the delayed missioners.

2. Data Understanding

This phase covers data collection and initial discovery of interesting subsets. As for the missions data, the IT staff in the Sector of Cultural Affairs and Missions, Ministry of Higher Education helped in the collection and understanding of the existing data, moreover this lead to the proper selection of data that can be used in the analysis process.

3. Data Preparation:

The data preparation phase includes the construction of the dataset for input into the modeling tool. Tables are selected, transformed and cleaned for the data modeling tool. Regarding the mission data, it was converted from flat files into relational tables in a relational database to start the analysis phase; these tables included the missioners data including the travel date and departure date and the extensions required for another year due to certain difficulties, it also includes the specialties and the countries the missioners went to.

4. Modeling:

A modeling technique is selected and applied in this phase. As for the missions case, a model was based on building a data warehouse containing the required fields for the entities that are to be investigated (missioners data, missions data, countries, specialties, departure and arrival dates and finally the extension request from the missioners).

The model was implemented without any external tools; SQL (structured Query Language) statements were executed against the database to retrieve the required information.

OLAP (On Line Analytical Processing) technique was implemented on the data warehouse. This technique has many advantages over the traditional statistical methods including the simplicity of data analysis for the business users, the powerfulness for data visualization, and the add on features like drilling up and down and pivoting to any level of the data levels.

5. Evaluation

At this stage the model is evaluated and reviewed. In order to make sure that all the parameters were carefully considered for the missioners data other parameters were then examined that affect the success of the missioners such as their academic back ground that is the university/institutions they are working in. Thus the model also includes the top universities/instititions and colleges that students works in before going on a mission.

6. Deployment

This includes organizing the data and the model in better ways for the customer use. For this phase, visual charts and reports were presented to the business users. The following section describes the information and the analysis results extracted from the data base analysis.

VIII.   ANALYSIS

Table (1)

The Missioners Achievements and their Marital Status

Missioners Achievements	Marital Status		Total	Percentage Married	Percentage Single	Percentage Missioners Achievements
	Married	Single				
Delayed	159	199	358	44%	56%	21.57%
Succeeded	777	525	1303	60%	40%	78.43%
Total	936	724	1661	56%	44%	100.00%

Table (2)

The percentage of delayed missioners in each country

Country	Delayed	Total	Percent Delayed
Austria	9	14	64.29%
Russia	8	13	61.54%
Sweden	6	11	54.55%
USA	97	313	30.99%
Holland	8	30	26.67%
Switzerland	2	8	25.00%
Germany	70	290	24.14%
France	24	102	23.53%
UK	82	391	20.97%
Denmark	1	6	16.67%
Spain	1	8	12.50%
Japan	25	202	12.38%
Italy	5	45	11.11%
Canada	18	186	9.68%
Belgium	1	14	7.14%
Australia	1	28	3.57%
Total	358	1661	21.55%

Table (3)

Missioners Specializations in Austria

Specialization	No of missioners	Percentage
Delayed missioners		
Basic Sciences	3	33.33%
Engineering	2	22.22%
Physical Education	1	11.11%
Agriculture	1	11.11%
Clinical medicine	1	11.11%
Arts	1	11.11%
Total	9	100.00%
Succeeded missioners		
Clinical Medicine	2	40.00%
Veterinary medicine	2	40.00%
Engineering	1	20.00%
Total	5	100.00%

Table (4)
Missioners Specializations in Russia

Specialization	No of missioners	Percentage
Delayed missioners		
Engineering	4	50.00%
Basic Sciences	2	25.00%
Agriculture	1	12.50%
Humanities	1	12.50%
Total	8	100.00%
Succeeded missioners		
Basic Sciences	2	40.00%
Agriculture	2	40.00%
Engineering	1	20.00%
Total	5	100.00%

Table (5)
Missioners Specializations in Sweden

Specialization	No of missioners	Percentage
Delayed missioners		
Pharmacology	3	50.00%
Commerce	2	33.33%
Basic Sciences	1	16.67%
Total	6	100.00%
Succeeded missioners		
Pharmacology	2	40.00%
Clinical Medicine	2	40.00%
Basic Medicine	1	20.00%
Total	5	100.00%

Table (6)
Marital Status of Delayed Missioners in Austria

Marital Status	Number of missioners	Percentage
Married	1	22%
Single	7	78%
Total	9	100.00%

Table (7)
Marital Status of Delayed Missioners in Russia

Marital Status	Number of missioners	Percentage
Married	1	13%
Single	7	88%
Total	8	100.00%

Table (8)
Marital Status of Delayed Missioners in Sweden

Marital Status	Number of missioners	Percentage
Married	2	33%
Single	4	67%
Grand Total	6	100.00%

Figure (1)
A graph showing the percentages of delayed missioners in each country
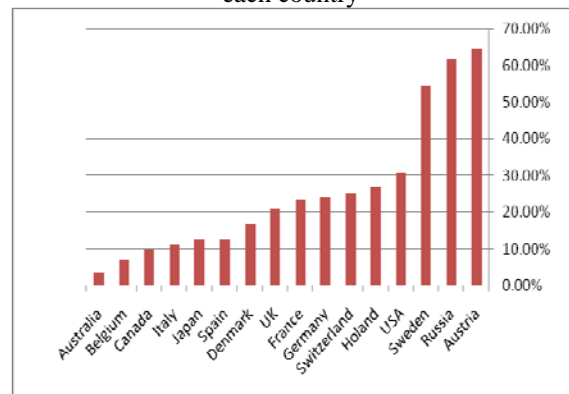
Table (9)
The percentage of succeeded missioners in each country

Country	Succeeded	Total	Percent Succeeded
Australia	27	28	96.43%
Belgium	13	14	92.86%
Canada	168	186	90.32%
Italy	40	45	88.89%
Japan	177	202	87.62%
Spain	7	8	87.50%
Denmark	5	6	83.33%
UK	309	391	79.03%
France	78	102	76.47%
Germany	220	290	75.86%
Switzerland	6	8	75.00%

Holland	22	30	73.33%
USA	216	313	69.01%
Sweden	5	11	45.45%
Russia	5	13	38.46%
Austria	5	14	35.71%
Total	1303	1661	78.45%

Table (10)
Missioners Specializations in Australia

Specialization	No of missioners	percentage
Succeeded missioners		
Basic Sciences	7	25.00%
Pharmacology	6	21.43%
Basic medicine	3	10.71%
Clinical medicine	3	10.71%
Engineering	3	10.71%
Agriculture	2	7.14%
Commerce	1	3.57%
Hotels and Tourism	1	3.57%
Humanities	1	3.57%
Total	27	100.00%
Delayed missioners		
Veterinary medicine	1	100.00%
Total	1	100.00%

Table (11)
Missioners Specializations in Belgium

Specialization	No of missioners	percentage
Succeeded missioners		
Basic Sciences	3	28.57%
Pharmacology	3	21.43%
Clinical medicine	3	21.43%
Basic medicine	2	14.29%
Engineering	2	14.29%
Total	13	100.00%
Delayed missioners		
Basic Sciences	1	100.00%
Total	1	100.00%

Table (12)
Missioners Specializations in Canada

Specialization	No of missioners	percent
Succeeded missioners		
Engineering	59	36.02%
Basic Sciences	28	15.59%
Basic medicine	21	11.83%
Agriculture	15	9.68%
Pharmacology	15	8.06%
Clinical medicine	14	7.53%
Veterinary medicine	6	4.30%
Humanities	4	3.23%
Commerce	2	1.08%
Law	2	1.08%
Physical Education	1	0.54%
nursery	1	1.8%
Grand Total	168	100.00%
Delayed missioners		
Engineering	8	44.44%
Basic Sciences	5	27.78%
Agriculture	3	16.67%
Hotel & Tourism	1	5.56%
Basic Medicine	1	5.56%
Total	18	100.00%

Table (13)
Marital Status of succeeded Missioners in Australia

Marital Status	Number of missioners	Percentage
Married	21	77.78%
Single	6	22.22%
Total	27	100.00%

Table (14)
Marital Status of succeeded Missioners in Belgium

Marital Status	Number of missioners	Percentage
Married	**8**	**61.54%**
Single	**5**	**38.46%**
Total	**13**	**100.00%**

Table (15)
Marital Status of succeeded Missioners in Canada

Marital Status	Number of missioners	Percentage
Married	**127**	**75.6%**
Single	**41**	**24.4%**
Total	**168**	**100.00%**

Table (16)
The top twenty Universities/Institutions with highest missioners success rates

University/Institution	Succeeded	Grand Total	percent succeeded
Center of Agricultural Research	3	3	100%
The Research Center for Housing and Construction	2	2	100%
Desert Research Center	2	2	100%
National Center for Social and Criminological Research	1	1	100%
Nuclear Materials Authority	1	1	100%
Atomic Energy Authority	1	1	100%
National Research Center	35	37	95%
Asuit University	139	161	86%
Ain Shams University	114	136	84%
Research centers of the Ministry of Scientific Research	49	59	83%
Tanta University	94	114	82%
Institutes of the Ministry of Higher Education	13	16	81%
Suez Canal University	117	145	81%
Zagazig University	63	80	79%
Azhar University	80	102	78%
Cairo University	86	111	77%
Mansoura University	95	124	77%
South Valley University	61	81	75%
Alexandria University	89	121	74%
Menia University	80	109	73%

Table (17)
The top ten colleges with highest missioners success rates

College / University the college is in	succeeded missioners	Total missioners	Percent succeeded
Medicine/ Tanta	28	28	100%
Pharmacy / Asyut	18	18	100%
Veterinary Medicine / Zagazig	13	13	100%
Dentistry / Suez Canal	11	11	100%
Languages / Menia	9	9	100%
Agriculture / Ain Shams	9	9	100%
Computers and information / Ain Shams	8	8	100%
Law / Asuit	7	7	100%
Science /Kina	7	7	100%
Arts / Souhag	6	6	100%

IX.   RESULTS

From the analysis of the higher education data for 10 years, it was found that the percentage of delayed and succeeded missioners is 22% and 78% respectively. As shown in Table (1) 44% of the delayed missioners were married and 56% were single while 60% of the succeeded missioners were married and 40% were single. The three countries with highest students delay rates are Austria (64%), Russia (62%) and Sweden (55%); this can be illustrated in Table (2). An analysis was done for these countries by drilling down for other criteria like the specializations studied at the universities and the marital status of the students. The drill down presented that the specializations of more than 50% of delayed students in Austria are related to basic sciences and engineering as shown in Table (3). Similarly, the specialization of 50% of delayed students in Russia and Sweden is engineering and pharmacology as clear in Tables (4) and (5). The delay of students is affected by their marital status as the percentage of single delayed students in Austria, Russia and Sweden is 78%, 88% and 67% respectively as presented in tables (6) to (8).

On the other hand, the three countries with highest student success rates are Australia (96%), Belgium (93%) and Canada (90%); this can be illustrated in Table (9). The specializations of 68% of succeeded students in Australia are related to basic sciences, pharmacology and medicine as shown in Table (10). Similarly the specialization of more than 50% of succeeded students in Belgium and Canada are basic sciences and pharmacology; and engineering and basic sciences respectively as presented in Tables (11) and (12). The success of students is affected by their marital status as the percentage of married students in Australia, Belgium and Canada is 78%, 62% and 76% respectively as presented in tables (13) to (15).

Moreover Table (16) reveals, the research centers that missioners work in and have 100% success rates are the Center of Agricultural Research, the Research Center for Housing and Construction, the Desert Research Center, and the National Center for Social and Criminological Research. The authorities that missioners work in and have 100% success rates are the Nuclear Materials Authority and the Atomic Energy Authority. The top three Universities that have the highest success rates are Asuit, Ain Shams, and Tanta with success rates of 86%, 84% and 82% respectively. Finally the top ten best colleges are presented in Table (17).

X.   CONCLUSION

The following section is the summarization and the conclusion for the analysis done on the ten years data related to the higher education missioners database:

The data showed that the majority of delayed students are single and the majority of successful students are married. This is also indicated by the single majority of delayed students in the three countries with highest student delay rate and the married majority of successful students in the three countries with highest student success rate. Even though Austria and Sweden have high delay rates yet missioners succeed in certain specializations which are veterinary medicine in Austria and clinical and basic medicine in Sweden. Moreover even though Australia, Belgium and Canada have high success rates, certain specializations

should be avoided. For example, students specializing in Veterinary medicine may be delayed in Australia. Furthermore 20% of students specializing in basic sciences are delayed in Belgium. As for Canada, 12% of students specializing in Engineering, 15% of students specializing in basic sciences, 17% of students specializing in agriculture and 100% of students specializing in hotel and Tourism are delayed. The knowledge of the best and worst specializations for each country may help in achieving effective decision making in the reallocation of Egyptian students to those countries

The study ranked the Universities/institutions and colleges with highest success rates. Further studies may compare Universities/institutions based on the University the missioners is working in and the success rate of each specialization that the missioners of this university perused on their mission.

REFERENCES

[1] D. Krbec, " Shaping New Paradigms in the Higher Education Development: Dilemmas for Transitional Countries," Informing Science, June 2002. Publisher@InformingScience.org

[2] E. N. Ogor, "Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques" Electronics, Robotics and Automotive Mechanics Conference, CERMA 2007, Page(s):354 – 359, 25-28 Sept. 2007.

[3] G. Abou Ali, " "An Integral Statistical and Analytical study of The Missions Central Administration during the period 2002-2008", Missions Central Administration , General Administration for Cultural Research. Sector of Cultural Affairs and Missions, Ministry of Higher Education, Cairo, Egypt 2008. (Arabic working paper)

[4] J. Luan, T. Kumar, T. R. Bohannon, P. M. James, B. L. Bailey, P. W. Eykamp and S. Sujitparapitaya, 'Data Mining Concepts, Myths and Case Studies, Mission: Effectiveness Through Diversity", 2006 Annual Forum, Association for Institutional Research, May 2006 Chicago, IL. http://www.airweb.org

[5] J. Luan, "Data Mining and Knowledge Management in Higher Education -Potential Applications", 42nd Annual AIR Forum, Toronto, Canada, June 2 - 5, 2002.

[6] J. Luan, M. Zhai, J. H. Chen, T. K. C. Chow, L. Chang and C.-M. Zhao," Concepts, Myths, and Case Studies of Data Mining in Higher Education", AIR 44th Forum Boston, May 30 – June 2, 2004.

[7] L. Daniel, "Discovering Knowledge in Data: An Introduction to Data Mining", Wiley, John & Sons, Incorporated, ISBN-13: 9780471666578, January 2005

[8] M. Zdravko, Daniel T. Larose, "Data mining the Web, Uncovering patterns in web content, structure and usage", Wiley-interscience, A John Wily & Sons,Inc., Publication, 2007

[9] N. Delavari, M. R. Beikzadeh, M. R. A. Shirazi, "A New Model for Using Data Mining in Higher Educational System", 5th International Conference on Information Technology based Higher Education and Training: ITEHT '04, Istanbul, Turkey, 31st May-2nd Jun 2004.

[10] N. Delavari, M. R. Beikzadeh, M. R. A. Shirazi, "Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System", 6th Annual International Conference on Information Technology based Higher Education and Training: ITEHT '05, Juan Dolio, Dominican Republic, July 7 – 9, 2005.

[11] N. Selmoune and Z. Alimazighi, " A decisional tool for quality improvement in higher Education" , Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on, 7-11 April 2008.

[12] P. Ramasubramanian, K.Iyakutti, V.Sureshkumar and P.Thangavelu, " Mining Analysis of SIS database using Rough Set Theory," International Conference on Computational Intelligence and Multimedia Applications, ICCIMA, IEE computer society, pages 81-87, 13-15 Dec. 2007.

[13] S. Colin, "Journal of Data warehousing", the data warehousing institute, Volume 5 Number 4, 2000.

[14] S. Wongpun1 and A. Srivihok, "Comparison of Attribute Selection Techniques and Algorithms in Classifying Bad Behaviors of Vocational Education Students," 2008 Second IEEE International Conference on Digital Ecosystems and Technologies, IEEE DEST 2008, Phitsanulok, Thailand, 26th – 29th February 2008.

[15] S. Z. Erdogan and M. Timor," A Data Mining Application in a Student Database", Journal of Aeronautics and Space Technologies, July 2005, volume 2, Number 2, (53-57), JULY 2005.

[16] V. P. Bresfelean, "Analysis and Predictions on Students' Behavior Using Decision Trees in Weka Environment," Proceedings of the ITI 2007 29th Int. Conf. on Information Technology Interfaces, Cavtat, Croatia, Page(s): 51-56, June 25-28, 2007.

[17] V. P. Bresfelean, M. Bresfelean,and N. Ghisoiu, "Determining Students' Academic Failure Profile Founded on Data Mining Methods", Proceedings of the ITI 2008 30th Int. Conf. on Information Technology Interfaces, Cavtat, Croatia, June 23-26, 2008.