An Approach for Integrating Data Mining with Saudi Universities Database Systems: Case Study

Mohamed Osman Hegazi[1]

Department of Computer Science
College of Computer Engineering
and Sciences
Prince Sattam Bin Abdulaziz
University
Al-Kharj, Saudi Arabia

Mohammad Alhawarat[2]

Department of Computer Science
College of Computer Engineering
and Sciences
Prince Sattam Bin Abdulaziz
University
Al-Kharj, Saudi Arabia

Anwer Hilal[3]

Department of Computer Science
Preparatory Year Deanship
Prince Sattam Bin Abdulaziz
University
Al-Kharj, Saudi Arabia

Abstract—This paper presents an approach for integrating data mining algorithms within Saudi university's database system, viz., Prince Sattam Bin Abdulaziz University (PSAU) as a case study. The approach based on a bottom-up methodology; it starts by providing a data mining application that represents a solution to one of the problems that face Saudi Universities' systems. After that, it integrates and implements the solution inside the university's database system. This process is then repeated to enhance the university system by providing data mining tools that help different parties -especially decision makers- to carry out certain decision. The paper presents a case study that includes analyzing and predicting the student withdrawal from courses at PSAU using association rule mining, neural networks and decision trees. Then it provides a conceptual and practical approach for integrating the resulted application within the university's database system. The experiment improves that this approach can be used as a framework for integrating data mining techniques within Saudi university's database systems. The paper concluded that mining universities' data can be applied as a computer system (intelligent university's system), Also, data mining algorithms can be adapted with any database system regardless that this system is new, exists or legacy. Moreover, data mining algorithms can be a solution for some educational problems, in addition to providing information for decision makers and users.

Keywords—*Data Mining; Database; Predict; Integration; Association rule mining; Neural networks; Decision tree; Educational Data Mining (EDM); University system*

I. INTRODUCTION

Nowadays, higher educational institutes face many challenges in fulfilling the needs and requirements of stake holders such as students and their parents, the public sector, the labor market and the community, besides the challenges of what is called the "revolution of academic accreditation". In additional to these challenges, the universities should keep pace with traditional issues such as predicting student enrollment and graduation rates, and the challenges of managing educational process to motivate higher educational institutes to search for better solutions. One way to effectively address these challenges is to use data mining technologies. Data mining has become one of the most valuable tools for extracting and manipulating data. It also used for establishing patterns in order to produce useful information for decision-making. Recently, data mining is used in educational field to bring a new field of study called Educational Data Mining (EDM). EDM focuses on collecting, archiving, and analyzing of data related to student learning and assessment in educational institutes. Through EDM, a university could, for example, predict academic outcomes, predict which students will or will not withdraw courses, predict which students will or will not graduate…etc. Also, using EDM we can get answers for many questions such as: what do institutions know about their students? Which courses have most withdrawal rate? What types of courses will attract more students? ...etc.

Recently, the usage of data mining in higher education has received more attention, although this doesn't apply to Saudi universities. (To the best of our knowledge, there is no use of data mining technologies in any Saudi university's educational system).

This paper tries to introduce data mining application to educational systems in Saudi universities; it provides an approach for using data mining algorithms in solving typical educational university problems and then propose an approach to integrate the proposed solution with Saudi university's database system.

There is a large number of research, experiments and projects in the area of EDM. For example:[1],[2], [3] and[4] work on the area of analyzing and predicting the course outcomes such as; classifying the student's success in one course, predict passing/failing rates in a course, predict withdrawal rates in a course and predict student's score in a course, while [5], [6], [7] and [8] work on the area of analyzing the performance of the academic and educational environment.

Also,[9],[10], [11] and [12] work on the area of analyzing, classifying and predicting academic success for students in a certain university level. In addition to predicting student performance, her/his graduation time, drop-out and general performance are analyzed, while in [13] the authors work in the area of examination and assessment , tackle predicting student's success in the next exam given her/his answers of previous exam. Also, they address predicting student's score in the next exam, and correctness of the student's answer.

In [14] and [15] the authors work on the area of metacognitive skills, habits, and motivation such as: predicting

the student's motivation or engagement level, his cognitive style, his experience in using the learning system, and recommended intervention strategy.

This paper contributes to the field of EDM by presenting a methodology to carry out the following:

1) Adapting data mining algorithms with database systems.

2) Providing an approach that might be used as a framework for integrating data mining techniques with Saudi universities' database systems.

3) Providing solutions (prediction and analysis) to one of the problems that is facing most Saudi universities (courses withdrawal).

4) Developing an EDM application.

The rest of this paper is organized as follows: section 2 presents the proposed approach, section 3 presents the experiments, and in section 4 results and conclusions are deduced.

II. THE PROPOSED APPROACH

This section describes the process of integrating data mining techniques with Saudi university's database system through a practical experiment. This is done based on a methodology that works in two phases (figure 1):

Phase one: (solving the problem)

In this phase, the problem (task) is selected and then developing a data mining application as a solution for the selected problem. This process includes the following steps:

1) Specify the problem.

2) Analyze the problem.

3) Specify and collect the dataset that is required for mining this task.

4) Prepare and preprocess the dataset (D1)

5) Develop the application.

Phase two: (integrate the solution)

This phase for integrating the application that is developed in phase one with the university's database system. The process of integration is achieved by the following steps:

1) Select the needed data for university database.

2) Add the selected data to application dataset (D1); producing in the new dataset (D2).

3) Update the data mining application to deal with the new dataset, resulting in the final data mining application.

4) Embed the application inside the university system (this can be done by adding the application's user interface and the reports inside the university user interface system – unifying the system).
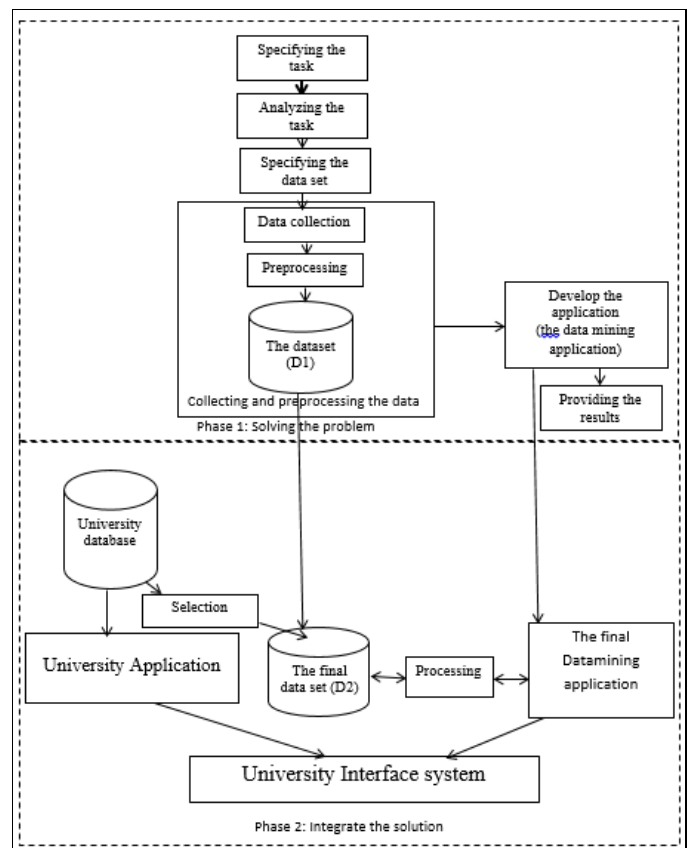

Fig. 1. The approach of integrating data mining techniques with the university's database system

This methodology can be used as an approach for embedding any type of data mining algorithm with Saudi university's database system. The idea of this approach is to select the required tasks and then apply the aforementioned steps for each task, continually. This concept, allow this approach -besides the integration of data mining techniques with university's systems- to provide two benefits:

- Well-defined life cycle for applying and integrating data mining techniques in university's database systems for certain tasks (figure 2).

- Flexibility in using and implementing any data mining technique in university's systems. Hence, it isolates the implementation of data mining technique (the data mining application) from the running university system. This is because that the approach provides the integration through the user interface without the need to implement the algorithms within the university application. In addition, the dataset that is used by the data mining technique is also isolated from the university database.
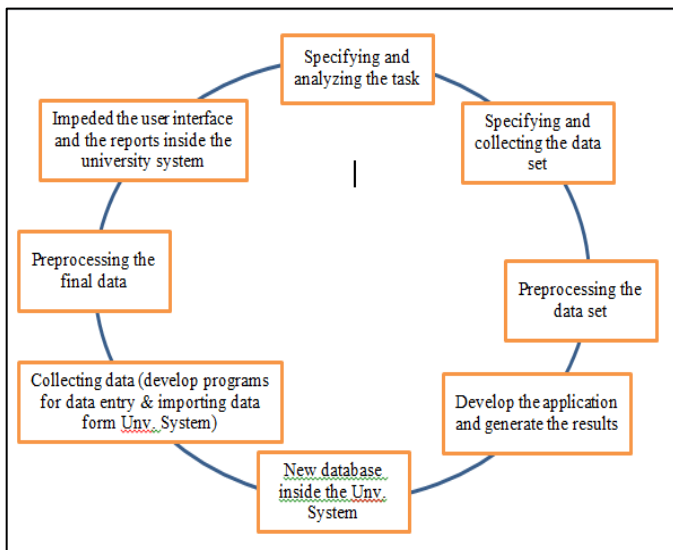
Fig. 2.   The life cycle for integration data mining techniques with university's database system

III.   THE EXPERIMENT

This section presents the experiment of the suggested approach. The experiment will be conducted at Prince Sattam Bin Abdulaziz University as case study of Saudi public universities.

A. Developing the datamining application (phase 1):

1) Specifying and analyzing the task:

Most of Saudi Universities allow students to withdraw from courses within 12 weeks after registration, bearing in mind that there are no fees for course registration or course withdrawal. Therefore, in every semester there are a considerable number of students who withdraw from one or more courses. Accordingly, withdrawal from courses constitute an academic burden and it is negatively affecting the educational process by: delaying student's graduation, affecting the performance of the programs, and sometimes affecting the following-up of the academic program specially when offering different optional courses.

The experiment will examine, analyze, and predict the student course withdrawal by applying association rule mining, neural networks and decision trees algorithms. The experiment integrates these algorithms in one application and then directly provides results that can support decision making and also provides a user interface that advises and helps students during course registration to avoid withdrawal.

In this experiment the implementation will be done inside the College of Computer Engineering and Sciences at Prince Sattam Bin Abdulaziz University, as specific case study. The final application can then be embedded into the university system, and can then be applied for all university colleges. The developing of the final application follows the algorithm that is shown in figure 1.

College of Computer Engineering and Sciences offers four programs, each program offers many courses, every student must enroll in one program but according to the program's

curriculum, students have to take some courses that are offered by other departments.

2) The dataset:

The most important data for analyzing and predicting student withdrawal from courses is reasons of the withdrawal (table 1). Unfortunately, such kind of data is not available in the university database. Accordingly, we surveyed 135 students to collect possible reasons for withdrawal from courses, 135 questionnaires have been distributed and filled by the selected students (students who withdrew courses in previous semesters). The responses of students to the closed-ended questions (mainly related to the reasons for withdrawal from courses) are used to build a database table to serve as a dataset for our experiment. Table 2 shows an example of the final records that preprocessed from the questionnaires.

TABLE I.   EXAMPLES OF REASONS OF WITHDRAWAL FROM COURSE AT PSAU

1.	The instructor (teacher): student disagree with or unlike the instructor
2.	Changing of the instructor (teacher): some time, for some reasons, the department changing course's instructor after registration
3.	Absent : student did not attends all the classes, so he may afraid to be denied or he found difficult in understanding or in following-up the course.
4.	Subject difficulty: student found that some or most of course subjects are difficult.
5.	Bad results (low marks): student found that his performance in the course work is not good (he obtain low marks in the course work), so he afraid to fail or he did not agree with his results.
6.	Option course: Because student has many choices, he may withdrawal from the option course to select another one next semester.
7.	Course timetable: Student may found that the course timetable is not suitable (he struggling in attending the classes)

the reasons has been specified according to results of direct interview with samples of students and academic advisors

TABLE II.   EXAMPLE OF THE QUESTIONNAIRE DATA AFTER PREPROCESSING (THE REASON OF COURSE WITHDRAWAL AND ITS RELATION WITH THE DEPARTMENT AND THE SUBJECT)

#	Subject	Sub. Tea.	Tea. Cha nge	Abs ence	Sub. Diff.	Sub. Time	Low Ma rks
1	Computer Network Systems	Yes	No	No	Yes	No	No
2	Computer Network Systems	Yes	No	No	Yes	No	No
3	Mobile Application	Yes	No	Yes	No	Yes	Yes
5	Discrete Mathematics	No	No	Yes	No	No	No
6	Logic Lab	Yes	No	No	No	No	No
8	Computer Network Systems	Yes	No	No	Yes	No	No
9	Web Applications Programming	No	No	No	Yes	No	No
10	Operating system	Yes	No	Yes	No	No	No

3) The application:

The application is designed to analyze and predict student's withdrawal through single user interface. In this application three data mining algorithms are used: Neural Networks (figure 3), Association Rule Mining (figure 4) and Decision Trees (figure 5). These algorithms is used to ensure that the application has the ability of applying different algorithms in

predicting student's withdrawal. SQL Server data mining because it is compatible with the University's database system and it has the ability to build multiple models on a single mining structure.
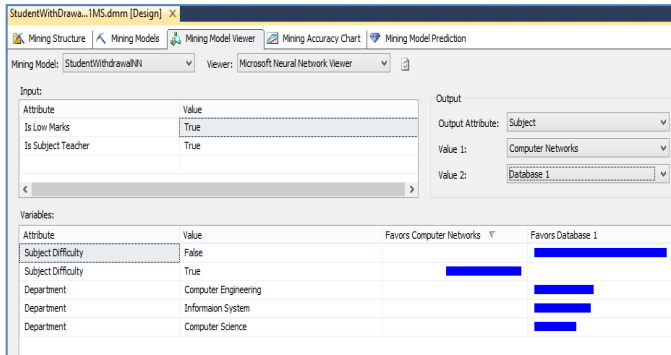

Fig. 3. The implementation of the neural networks data mining algorithm (Student withdrawal data mining application)
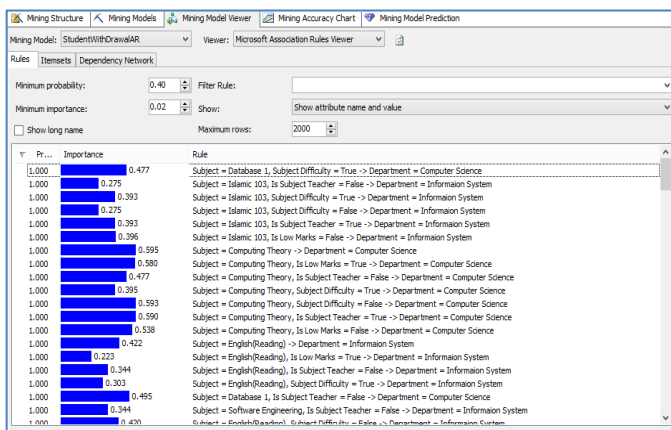

Fig. 4. The implementation of the association rule mining algorithm (Student withdrawal data mining application)
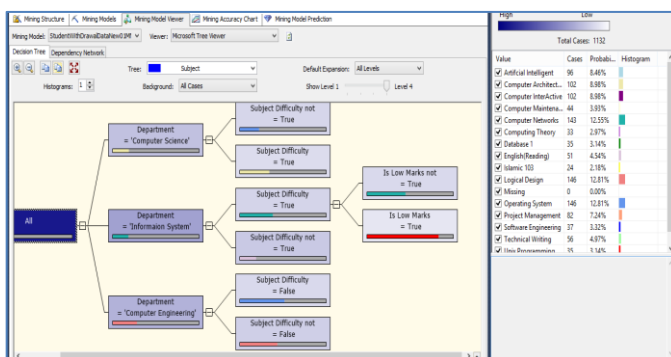

Fig. 5. The implementation of the decision trees algorithm (Student withdrawal data mining application)

These algorithms are integrated into one user interface in which the user can then choose the algorithm, subject, confidence, … etc. Then, results appear providing the prediction for the student's withdrawal from courses according to the selected mining algorithm (figure 6).
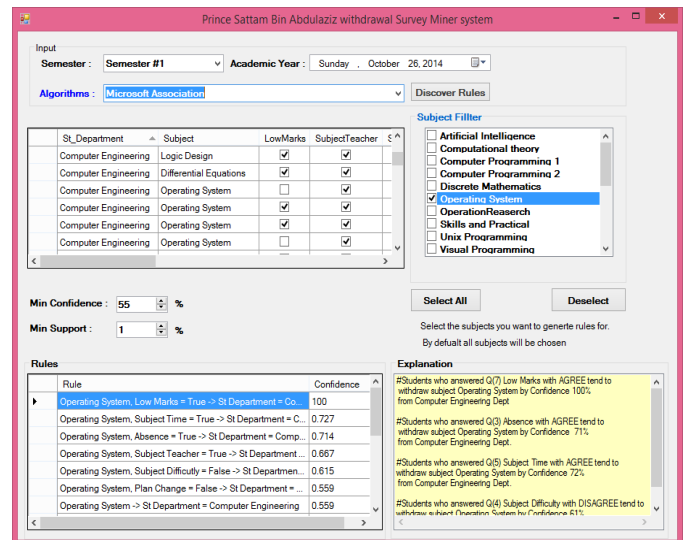
Fig. 6. Student withdrawal data mining application (main interface)

B. Integrating the solution (phase 2)

1) The final dataset

The final database is designed to handle data coming from two datasets:

a) The first dataset is about reasons of withdrawal. This data has been created and developed in the previous phases (see figure 1 – D1).

b) The second dataset is for the information about students, courses, and instructors. This kind of data is selected directly from the university database.

These two datasets are then joined in one database to present the final dataset for the final data mining application (see figure 1).

2) The final datamining application:

In this stage the final data mining application is developed. The application is designed and developed to be an automatic application (computer system) that can run automatically inside the university system without any manual processing. To provide this concept, the final datamining application (the mining system) is designed to perform the mining through three stages:

- At the first stage, the application constructs the classifier (figure 7).

- At the second stage, the application uses the classifier for classification (figure 8).

- Al last, all the processes are provided through an online application (figure 9).

The first step: constructing the classifier:

In this step, the classification algorithm constructs the classifier from the training set made up of database records and their related class name coming from the final dataset (D2). Figure 7 explains the construction of the classifier.
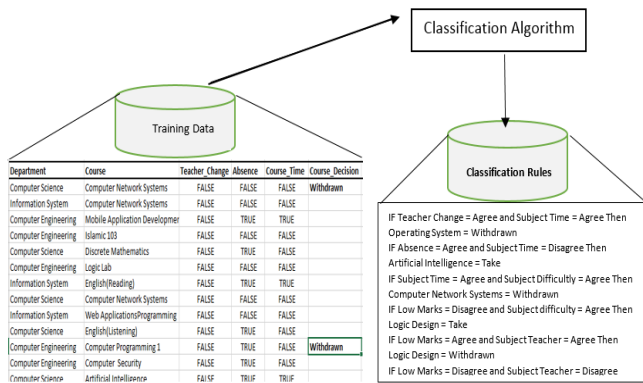
10 show the results of predicting the withdrawal of three courses (Artificial Intelligence, Operating Systems and Human Computer Interaction). These results are based on three classification models built automatically when the user select the required items and press the appropriate buttons. The system gives a confidence percentage for this prediction by comparing predicted percentage with set minimum confidence.

Fig. 7. Classification algorithms

The second step: The classification:

In this step, the classifier is used for classification. Here the test data is used to estimate the accuracy of classification rules. Figure 8 explains the use of the classifier for classification.
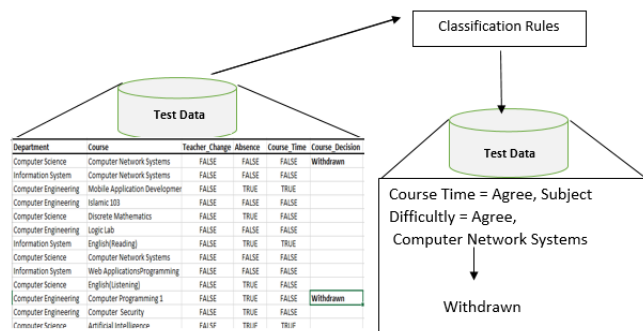
Fig. 8. Classification Rules

The last step: The online data mining application:

At the final stage, all these processes are provided online through a single user interface form (figure 9). The user can then interact with the mining system through the online system's interface by selecting the corresponding items (figure 9).
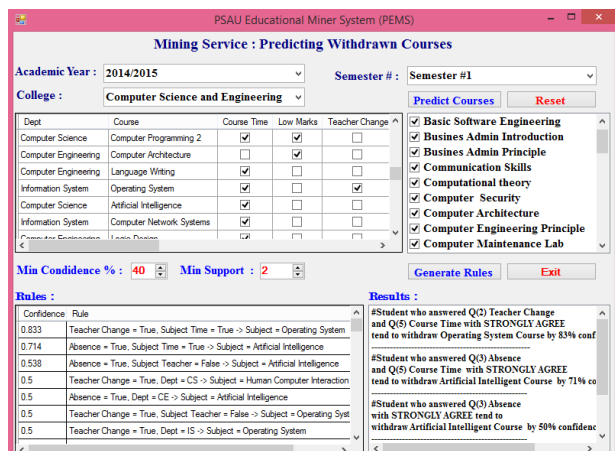
Fig. 9. PSAU Educational Miner System interface

The system provides predicting results based on the mining algorithm that is selected by the user, for example figures 9 and

Fig. 10. Samples of some predicted results

The final mining system:

The final mining system that is provided by this experiment is now tested and ready to run within the university system. The student and the decision maker can make use of this system. For example, student can choose the right course to avoid any future needs for withdrawal, the decision maker can get a lot of information that can help in developing the right action plan regarding the processing of the teaching and the registration processes.

At any semester, student can withdraw courses, which means new withdrawal data is arising continually, this kind of data is most close to reality; hence, it is affected by the current situation of the registration process. Accordingly, we make use of this data by providing new interface form to collect these kind of data instead of using the questionnaire, to keep the automaticity of the system, provide the continuation of the system, and to let the system be more intelligent by adding the new data to the old dataset (learning knowledge base concept) (figure 11). When the student decides to withdraw a course, then he will be asked through this form to enter the reasons of withdrawal, his answers then are added to our database (dataset D2). Accordingly, the users deal with the system as follow:

a) At the registration period, the student can check the probability of the withdrawal of the courses he is going to register, the reasons that he may face during the semester, so he will make his right decision. For example, if he afraid from the time, or the instructor of the course he then asks the system what is the possibility of withdrawal of this course because of the time or the instructor.

b) The administrators can make the right action plan. For example they can solve or avoid the reasons that may affect the withdrawal rates when offering courses at the beginning of each semester.

c) When the student decides to withdraw from course(s) he/she will be asked to enter the reasons of withdrawal (figure 11) this data is then added automatically to the system dataset. (And so on, the cycle will continue starting from 1 or 2 above, according to who will use the system – student or the decision maker).

Fig. 11. Reasons of withdrawal (data entry form)

IV. RESULTS AND CONCLUTIONS

The paper presents an approach for embedding data mining algorithms with Saudi university's database systems. The approach is based on a concept that embeds data mining algorithms with the running university system by developing a separate application and separate dataset, and then integrates all the processes with the university system through unifying the application user interface with the university system user interface. This concept provides more flexibility in using and implementing any data mining technique in university systems.

The approach is tested at Prince Sattam University (one of Saudi public universities). The experiment shows that a data mining algorithm can be embedded into the Saudi university's system and would provide intelligent results that might help students, parents and decision makers. In addition, the experiment shows that data mining algorithms can be implemented as a computer system running automatically in a continuous manner without any intervention or manual preprocessing.

The experiment concludes that the approach can be used as a framework for integrating data mining techniques with Saudi universities' database systems, besides providing solutions (prediction and analysis) to one of the problems that is facing most of Saudi universities: courses withdrawal.

This paper concludes that, mining universities data can be applied as a computer system (intelligent university system), data mining algorithms can be adapted with any database system whatever this system is new, exists or legacy, and data mining algorithms might provide solutions for some educational problems, in addition to providing information for decision makers and users.

REFERENCES

[1] Calders, Toon, and Mykola Pechenizkiy. "Introduction to the special section on educational data mining." ACM SIGKDD Explorations Newsletter 13.2 (2012): 3-6.

[2] Chacon, Fabio, Donald Spicer, and A. Valbuena. "Analytics in support of student retention and success." Research Bulletin 3 (2012): 1-9

[3] Hämäläinen, Wilhelmiina, and Mikko Vinni. "Classifiers for educational data mining." Handbook of Educational Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series (2010): 57-71.

[4] Yamamoto, Yukiko, et al. "Increasing the Sensitivity of a Personalized Educational Data Mining Method for Curriculum Composition." Emerging Issues in Smart Learning. Springer Berlin Heidelberg, 2015. 201-208.

[5] Bienkowsk Marie i, Feng Mingyu and Means Barbara. 2012." Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief". Technical report. U.S. Department of Education. Office of Educational Technology. October 2012

[6] Delavari Naeimeh, Phon-Amnuaisuk Somnuk and Beikzadeh Mohammad Reza 2008. " Data Mining Application in Higher Learning Institutions" Informatics in Education, 2008, Vol. 7, No. 1, 31–54 31

[7] Titus deLaFayette Winters. 2006. "Educational Data Mining: Collection and Analysis of Score Matrices for Outcomes-Based Assessment". PhD. Thesis in Computer Science. UNIVERSITY OF CALIFORNIA RIVERSIDE

[8] Zare Ali Akbar, Gerayeli Mohamad, Zare Davoud, and Tadayon Hamed. 2013."Teachers Salary Data Analysis using Data Mining Techniques" European Journal of Economics, Finance and Administrative Sciences

[9] Arockiam, L., S. Charles, and M. Amala Jayanthi. "An Impact of Emotional Happiness and Personality in Students' Learning Environment." Data Mining and Knowledge Engineering 7.2 (2015): 69-74.

[10] Baradwaj, Brijesh Kumar, and Saurabh Pal. "Mining educational data to analyze students' performance." arXiv preprint arXiv:1201.3417 (2012).

[11] Chalaris, Manolis, et al. "Examining students' graduation issues using data mining techniques-The case of TEI of Athens." Proceedings of the 4th International Conference on Integrated Information. Vol. 1644. AIP Publishing, 2015.

[12] Patidar, Preeti, Jitendra Dangra, and M. K. Rawar. "Decision Tree C4. 5 algorithm and its enhanced approach for Educational Data Mining." (2015).

[13] Foltz, Peter W., and Mark Rosenstein. "Analysis of a Large-Scale Formative Writing Assessment System with Automated Feedback." Proceedings of the Second (2015) ACM Conference on Learning@ Scale. ACM, 2015.

[14] Baker, Ryan Shaun, and Paul Salvador Inventado. "Educational data mining and learning analytics." Learning Analytics. Springer New York, 2014. 61-75.

[15] Ocumpaugh, Jaclyn, et al. "Population validity for Educational Data Mining models: A case study in affect detection." British Journal of Educational Technology 45.3 (2014): 487-501.