Volume 6, Issue 3, March 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Proposed Framework for Data Mining Techniques in Learning System

Ayman Khedr, Amany Abdo, Aya Amar
Faculty of Computer & Information. Helwan University,
Cairo, Egypt

Abstract*— Today, many universities seek to improve their efficiency and effectiveness of educational quality. In response to such a demand, this paper has three main contributions. First, it introduces a survey of the available literature according to different topics for data mining techniques in the educational environment .Moreover, discusses their advantages and drawbacks. Second, the paper includes a comparison study of the educational data mining algorithms based on some comparison factors that have been introduced. Finally, it provides a new framework and implementation processes for applying data mining (DM) techniques to enhance students' performance level in the higher education. A framework supports decision making based on the extracted information to take the right decision at the right time.

Keywords*— Data Mining; Learning System; Decision Making.

I. INTRODUCTION

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge [5]. Data mining are organized according to the type of modelling techniques used, which include: Neural Networks, Genetic Algorithms, Clustering and Visualization Methods, Fuzzy Logic, decision tree, support vector machines, intelligent agents, and Inductive Reasoning, amongst others. Knowledge discovery in databases (KDD), often called data mining aims at the discovery of useful information from large collections of data [9][16].

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational system, and using those methods to better understand students and in the same time it achieve these benefits as: minimizing wastage education through selecting the appropriate decision for students to increase education values[12].

The educational data mining has been classified as:

- Prediction (Classification, Regression, Density estimation)
- Clustering
- Relationship mining (Association rule mining, Correlation mining,
- Sequential pattern mining, Causal data mining)
- Distillation of data for human judgment
- Discovery with models [13].

The process of creating vital information from a huge amount of data is learning. It can be classified into two such as supervised learning and unsupervised learning [6].Today, one of the biggest challenges that educational institutions face is the explosive growth of educational data and to use this data to improve the quality of managerial decisions. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from large data sets [7].

Vranić, M., et al , examined how to improve some aspects of educational quality with data mining algorithms and techniques by taking specific course students as target audience in academic environments [16]. Another study on data mining applied in education was published in 2000 by Becker and others which are performed for defining and understanding the impact of changes in curriculum on students at a university in Brasil [3].

A data mining application in which defining of student characteristics are used for measuring the satisfaction of students at higher education was performed by Luan in 2002 [8] .Several data mining tools which assist to extract knowledge and discover patterns can be used. Such tools can be commercial or open source. Example of commercial data

mining tools SAS Enterprise Miner, SPSS IBM Intelligent Miner , Microsoft SQL Server 2005 , Oracle Data Mining (from Oracle 10g) , Angoss Knowledge STUDIO , KXEN. Example of open source data mining tools R, Tanagra, Weka ,YALE, KNIME, Orange, GGobi [10][14] [18].

The research is structured as follows: Section two surveys data mining techniques in higher education and a comparison study according to different topics for education data mining techniques with highlighting the advantages and disadvantages of each research. Finally, Section three concludes the research and identify the future research.

II. RELATED WORK

Several studies have demonstrated that Data Mining techniques could successfully be incorporated into the learning process. So we briefly introduced some of valuable case studies. One of the first studies on data mining implemented in education was published in 1995 by Sanjeev and Zytkow. Researchers collected the knowledge discovery as terms like "P pattern for data in the range R" from university database [14].

Aher, S., et al, discussed the using of data mining techniques in the educational system. Moreover, the result analysis was shown with the help of WEKA tool. This approach tried to prove the usefulness of data mining in higher education specifically for improving the students' performance [1]. After studying this research, it has been found that the main advantage of this research was using DBSCAN algorithm which handled the outlier problem. However, some disadvantages have been occurred for example the sample size used was small as it contained a few number of students which led to inaccurate results. Additionally, the number of courses were small which may affect the dependency of these results.

Osmanbegović, E., et al, aimed to build a model to attain the conclusion on students' academic success. They used three supervised data mining algorithms which were applied on the preoperative assessment data in order to predict success in a course. The main purpose of this methodology was to support students and teachers to enhance student's performance. Moreover, it aimed to decreasing the failing ratio by taking convenient steps at the suitable time to enhance the quality of learning [11].

After our study to this research, it shows that there are some advantage for examples: The evaluation of results during two classes coded in this way category (A) failed, category (B) passed and this may led to a few prediction error rate. Navia Bayes algorithm outperformed in prediction of decision trees as it's consumed a slight time. However, the main disadvantage of this research is applying one course and this may lead to inaccurate outcomes.

Ahmadi, F., et al, introduced the usage of data mining in teacher evaluation system and also performed on outcome analysis using WEKA tool. They aimed to collect the manageable experiences with data mining also utilizing of these experiences at E-learning system and traditional education based on teacher evaluation. This methodology can summarize variety outcomes which help education managers in universities [2].

After studying this research, it has been found that the main advantage of this research was using a new trend for testing actor to evaluate teachers' performance. This help to predict which teachers will be invited to faculty classes and which teachers will be refused.

Tair, M. M. A., et al, adopted the use of educational data mining to enhance the graduate student's performance and also face the problem of low grades of graduate students .A case study in the educational data mining has been presented which showed the usefulness of data mining in higher education especially to enhance the performance of graduate students [15] .

After our study to this research, it shows that the main advantage of this research is using Bayes-naïve bayes and Rule induction-(if-then) which are can predict low grades on time. For example, the college management can predict Average students from the beginning and they may work on the students to improve their performance before graduation.

Yadav, S. K., et al, introduced a data mining project in order to yield predictive models for student retention management. The quality of these predictive models, generated by the machine learning algorithms, were tested in this research. The results of this methodology exhibited that we can make short and also accurate predictions list for the student retention purpose through implementing the predictive models to the records of new incoming students. Moreover, the research defined the students which need special care to reduce dropout rate [17].

After studying this research, it has been found that the main advantage of this research was using alternative decision tree (ADT) learning algorithm as it outperformed in predicting decision tree and also building the model more rapidly. However, the main disadvantage of this research is not discussing the different skills and Knowledge levels between students. However identifying these factors are very important to decrease drop-out rate and improving learning results.

The table below introduces a comparison study of the educational data mining algorithms based on some comparison factors. Table 1 includes the following information: Research title, Objective, Data Set, Testing Actor, Future Work, Software used, Data Gathering Method, Sample size, Technique, Algorithms, and Evaluations Results.

Table 1: A Comparison Study between Data Mining Techniques in Education Environment.

Research Title	Mining in Educational System using WEKA (2011) [1]	Data Mining Approach for Predicting Student Performance (2012) [11]	Data Mining in Teacher Evaluation System using WEKA(2013) [2]
Objective	To predict drop-out student, relationship between the student university entrance examination results and their success, predicting student's academic performance, discovery of strongly related subjects in the undergraduate syllabi, knowledge discovery on academic achievement, classification of students' performance in computer programming course according to learning style, investing the similarity and difference between schools.	To produce a model which would stand as a foundation for the development of decision support system in higher education.	To predict that which teachers will be invited to faculty classes and which teachers will be refusing.
Data Set	They collected all available data including their performance at university examination in **4** courses {**ACA (**advanced computer architecture) **MIS** (management information system) **ADS** (advanced database system) **OOMB** (object oriented modeling and design)}	**2** variables were used as inputs for Business information courses.(Gender, Distance, **GPA(*) ,**Scholarships, **Materials(*),** Family, High School, **Entrance exam(*), Time,** the Internet, Earnings ,Grade importance)**Notes: (*)** data set was the most important	**5** variables were used as input (Evaluation's score, Teacher's degree, Degree's Type,Teaching experience ,and Acceptation)
Testing Actor	GPA in various subject in information technology department. Two classes (Passed-Failed).	Students and courses	Teacher Performance in information science. Two classes (Yes-No)
Future Work	This study will be more efficient if it apply various courses to get more accurate results using more data mining techniques such as neural nets, genetic algorithms, k-nearest Neighbor, Naive Bayes, support vector machines and others.	This methodology can be used to help students and teachers to improve student's performance; reduce failing ratio by taking appropriate steps at right time to improve the quality of learning through answering these questions: How to obtain that predicting models are user friendly for professors or non-expert users? How to integrate data collection system of university and data mining tool?	The education managers could use these rules to take the best decision in the future to improve the students learning outcomes.
Software used	WEKA	WEKA	WEKA
Data Gathering Method	Students 'data from database of final year.	Questionnaire	Questionnaire
Sample size	85 student	257 Student	104 teacher 803 student
Technique	Association- Classification- Clustering	Classification. (4) tests: for the assessment of input variables: Chi-square test, One R-test Info Gain test, Gain Ratio test	Association Classification Cross validation fold Clustering
Algorithms	Rule- Zero R , DBSCAN	Tree –NB, Tree –MLP, Tree -J48/C4.5	Tree-j48, Rule- Zero R
Evaluations Results	DBSCAN algorithm handles the outlier problem.	There were **3** algorithms (NB, MLP, and J48/C4.5) that used to build model. **Criteria** were used to evaluate 3 models (execution time, and precision value). The results illustrated that (NB) Navia Bayes classifier outperforms in prediction decision tree. Because MLP consume more time to build the model but NB more rapidly in the time to build the model. NB: 0 / MLP: 4.13 / J48: 0 Time Precision: 76.65 / 71.2 / 73.93	Teacher's evaluation, evaluation's score of students is very important factor that many universities gather this information on performance of teacher.

III. A PROPOSED FRAMEWORK AND PROCESSES

A proposed framework for learning system contains three basic phases: The first phase is about capturing raw data a university or external environment. The second phase involves filtering educational data (selected, cleaned, and transformed students & courses data). In addition to, applying data mining techniques. The third phase includes selecting decision based on extracted knowledge value as shown in fig.1.

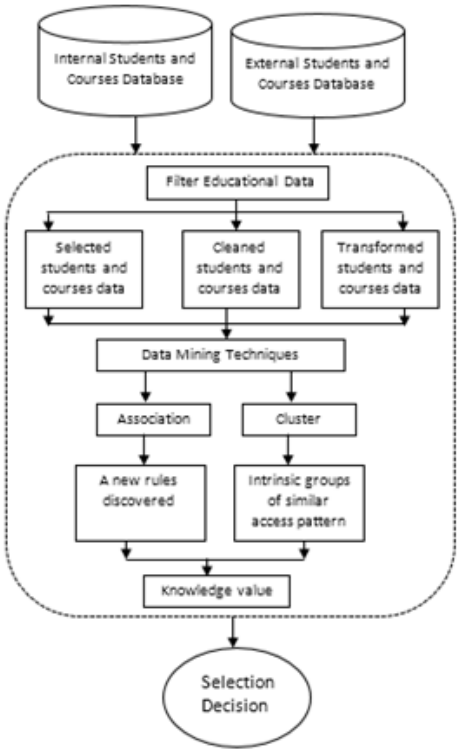

Fig. 1. A proposed framework for learning system

A proposed framework for learning system and implementation processes consists of five steps: The first step is about identifying student request. Furthermore, analyzing and understanding it .Step two includes collecting raw data from a university, filtering educational data. Moreover, applying data mining techniques to produce meaningful information. Step three involves identifying alternative decisions, such as: registering course, no course registration, registering in addition to introducing tips, and then chooses the best course. Step four implements the selected course. Step five measures the performance in order to, evaluate and determine whether a new framework is moving towards the achievement of its objective or not. If the student' performance level is not improved, return to step one or two as appropriate as shown in fig.2.

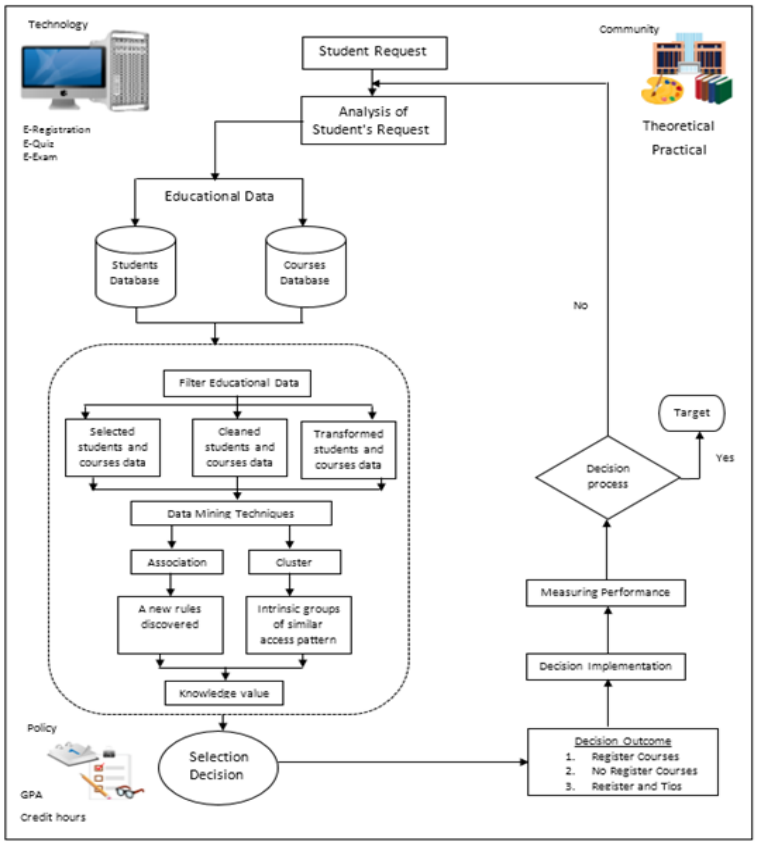

Fig. 2. A proposed framework for learning system and implementation processes

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we investigated in the problem of efficiency and effectiveness of educational quality. We reviewed the issues of students and teachers in the educational sector. A survey and a comparison study of the educational data mining algorithms have been accomplished. A proposed framework suggests courses for a student according to his skills regarding to new rules discovered. The framework presents useful advices and information for higher educational system to take more effective and efficient decisions at the right time. Due to the rapid growth of students' numbers however not all the students have the same mental level; there are genius student, high intelligent student, average intelligent student, low intelligent student, etc. Furthermore, there are no interactive guiding system to support students during academic study. Therefore, there is an urgent need to develop the general framework that could enhance the students' performance level. Moreover, student's evaluation and courses evaluation are very important factors that should be considered when applying the proposed framework

REFERENCES

[1] Aher, S. B. and L. Lobo (2011). Data mining in educational system using Weka. IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT) (3) pp. 20-25.

[2] Ahmadi, F. and S. Abadi (2013). Data Mining in Teacher Evaluation System using WEKA. International Journal of Computer Applications, Vol.63, No.10, pp.12-18.

[3] Becker, K., et al (2000). Using KDD to analyze the impact of curriculum revisions in a Brazilian university. AeroSense 2000, International Society for Optics and Photonics, pp.412-419.

[4] Castro, F., et al (2007). Applying data mining techniques to e-learning problems. Evolution of teaching and learning paradigms in intelligent environment, Springer,pp. 183-221.

[5] Han, J., et al (2006). Data mining: concepts and techniques, Elsevier.

[6] IndiraPriya, P. and D. Ghosh. (2013). A survey on different clustering algorithms in data mining technique.International Journal of Modern Engineering Research, Vol.3, No.1, pp.267-274.

[7] Kumar, V. and A. Chadha (2011). An empirical study of the applications of data mining techniques in higher education. International Journal of Advanced Computer Science and Applications Vol.2, No.3.

[8] Luan, J., (2002). Data Mining and Knowledge Management in Higher Education-Potential Applications.

[9] Mannila, H, 1996. Data mining: machine learning, statistics, and databases. Ssdbm, IEEE.

[10] Mikut, R. and M. Reischl (2011). Data mining tools. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol.1, No.5, pp.431-443.

[11] Osmanbegović, E. and M. Suljić (2012). Data mining approach for predicting student performance. Economic Review, Vol.10, No.1.

[12] Romero, C. and S. Ventura (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, Vol.33, No.1, pp.135-146.

[13] Romero, C., et al (2010). Class association rules mining from students' test data. Educational Data Mining 2010.

[14] Sanjeev, A. P. and J. M. Zytkow (1995). Discovering Enrollment Knowledge in University Databases. KDD, pp.246-251.

[15] Tair, M. M. A. and A. M. El-Halees (2012). Mining educational data to improve students' performance: a case study. International Journal of Information, Vol. 2, No.2.

[16] Vranić, M., et al (2007). The use of data mining in education environment. Telecommunications, 2007. ConTel 2007. 9th International Conference on, IEEE, pp.243-250.

[17] Yadav, S. K., et al (2012). Mining Education data to predict student's retention: a comparative study.arXiv preprint arXiv: 1203.2987.

[18] Zupan, B. and J. Demsar (2008). Open-source tools for data mining. Clinics in laboratory medicine, Vol.28, No.1, pp.37-54.

