Data Mining and Gamification Techniques in Adaptive E-Learning: Promises and Challenges

Reem S. Al-Towirgi[1,2]

Lamya F. Daghestani[1]

Lamiaa F. Ibrahim[1,3]

[1] Department of Computer Science
Faculty of Computing and Information Technology
King Abdulaziz University Jeddah, Saudi Arabia

[2] Department of Computer Science
Faculty of Computing and Information Technology
Taif University, Taif, Saudi Arabia

[3] Computer Science Department
Institute of Statistical Studies and Research, Cairo University
B.P. 42808 Zip Code 21551- Girl Section, Jeddah, Saudi Arabia

ABSTRACT
Educational Data Mining EDM is an emerging discipline. It concerned with extracting useful information from large educational data. It serves education improving by presenting information to facilitate the process of decision making. EDM has many methods and applications the context of e-learning. Gamification is the process of using mechanics and dynamics of games onto non-game context to promote the desired behavior. An emerging type of learning method is the adaptive e-learning. This paper discusses the state of the art of EDM and gamification methods to build adaptive e-learning system.

General Terms
E-learning, Data Mining, Gamification.

Keywords
E-learning, Learning Management System, Educational Data Mining, Knowledge Discovery in Databases, Gamification, Adaptive E-Learning.

1. INTRODUCTION
In the last few years, with the evolution of information and communication technologies (ICT), many domains were evolved. One of the most affected fields is education. By using ICT in education, e-learning was created [1].

E-learning systems keep a massive amount of data about students and their behavior in these systems. Also, e-learning systems contain data about instructors and courses [2]. So we need to advanced methods for analyzing these data and for extracting useful information and representing it in a way to facilitate the process of decision making to increase the effectiveness of learning process.

Data mining is used to achieve this goal through some techniques and methods [3]. Data mining or knowledge discovery in databases (KDD) is the process of extraction of interesting patterns from large dataset [4]. A special field of data mining that used in education context is called Educational Data Mining (EDM).

Games begin to pervade the real world and have been a large part of human life. It can be used in any industry to keep the users interactive and interested. The motivational power of games can be taken and applied to any real-world problem to get a gamified system [5]. Gamification is the process of using mechanics and dynamics of games onto non-game context to promote the desired behavior [6]. Games mechanics are the elements that move the actions forward such as points, bonuses, and levels. Game dynamics are the tools that help to figure out how to move the action forward such as achievements, competition, status, and altruism. Gamification is different from serious games that create full-fledged games to achieve the same goal. Rather than creating an actual game, gamification grafts elements that make video games fun and effective and applying them to applications.

This paper aims to present the state of the art of methods of EDM and gamification in e-learning systems. This paper is organized as follow: section two presents a background of EDM and its relation to e-learning. A review of studies used data mining in education and their applications in section three. Description of gamification techniques in adaptive e-learning will be in section four. Finally, a roadmap and conclusion are presented in section five and six respectively.

2. E-LEARNING AND EDUCATIONAL DATA MINING
E-learning is defined as training delivered in on a digital device to support individuals learning and achieve goals [1]. To facilitate the interaction between students and teachers in e-learning, Learning Management Systems (LMSs) was created. Nowadays, most of the institutions of higher education are using LMSs to facilitate the learning experience [7]. Examples of popular LMS are Blackboard, Moodle, Edmodo, and Canvas [8].

To support the learning process, LMSs provided different features. LMS features are categorized into four basic categories depending on what they support [9]: administration, content, assessment, and engagement. Administration features concerned with course administration aspects such as calendar and announcements. Content features support accessing the course modules and linking to external resources. Assessment features facilitate the evaluation

process through quizzes and assignments. Engagement features concerned with users participation such as discussion forums and collaboration activities. These features can be cooperated to enhance many different courses in different disciplines.

A unique feature of educational data is that its hierarchical structure. This means that data from different levels such as the school level, the teacher level, the classroom level, the student level, the session level, the answer level and the keystroke level are nested inside one another [10], [11]. Also, there are three important data attributes need to be considered while capturing educational data which are: time, sequence and context. Time parameter concerned with the time of learning and the length of sessions. Sequence concerned with the order of building content and practice. Context helps in explain results and decide the availability of a model working [12]

Most of the LMSs store the users' data when they interact with the system in a large data set for further analysis [13]. These data are different from the traditional educational dataset in terms of being tall (number of students), wide (attributes for each student), long (in duration).

LMSs differs in terms of stored information and how they storing this information. Some LMSs store the number of times each property is used for every user while others store the updates of features used by each student over the days. Also, some LMSs use a relational database to store data while others use flat log files [14].

Information about learner's navigation through the course and how instructors support the students learning through course design create massive datasets. Researchers tend to analyze these data sets using the computational and statistical methods emerging from the big data field to discover a new pattern of learning and teaching [15][16]. This leads to a new field which uses the applications of data mining methods to an educational dataset called educational data mining (EDM) [17].

Educational Data Mining (EDM) extends from Data Mining (DM) field which attempted to achieve many requirements that could not be achieved by traditional methods using statistics, artificial intelligence, pattern recognition and machine learning [11]. EDM results can be used to support learning and teaching in an innovative way through learning analytics [18].

The purpose of EDM is to extract useful information from large learning dataset collected from online learning platforms to improve learning [18]. It provides useful information to students, faculty, and administrators such as: predicting student success or failure, extracting valuable pattern and creating a model for student learning, and evaluating e-learning process, Examining the relationship between students characteristics and learning outcome, monitoring student persistence and informing instructional design [12].

3. EDUCATIONAL DATA MINING METHODS AND APPLICATIONS

Several methods and techniques are used in EDM researches. Baker and Yacef [10] suggest that EDM is used to improve student model, improve domain model and analyze the support of learning system. They categorized the EDM methods into prediction, clustering, association rule mining, a discovery with models and distillation of data for human judgment.

Romero and Ventura [11] found that the most DM methods used are regression, clustering, classification and association rule mining. They categorized the applications of EDM into a visualization of data, providing feedback to support teachers, a recommendation to students, predicting learner performance, modeling of students.

EDM can be used to analyze data from any interactive learning environment. In this paper, the focus will be on EDM researches that applied to higher education. These researches are presented in the following sections and categorized according to the method used.

3.1 Prediction
Prediction is the process of developing a model to infer what behavior will predict the student success or failure in the learning outcome [18]. Statistical methods are used in prediction studies such as multiple regression and logistic regression. LMS variables can be used to predict the performance of student such as the number of assignment completed, the number of quizzes passed, and the number of posts in discussion forum etc.

In [19], Macfadyen and Dawson analyzed LMS student usage of online course. They use correlation analysis, multiple regression analysis, and logistic regression to predict the final grade. They found that final grade was predicted depending on the total number of posted topics in the forum, total number of sent mail and the total number of completed assignment.

Abdous et al. in [20], study the relationship between the question type asked by a student in the forum and his final grade. In their paper, researchers classify the questions into four themes: class check-in, deadline/schedule, evaluation/technical and learning/comprehension. Results showed that students that asked about learning/comprehension had a higher final grade.

Thakur et al. in [21] used logistic regression and neural network model to predict student success. The number of completed assignments was the basic predictor of student success and failure, followed by the completed quizzes and submitted posts.

In [22] Jo et al. used multiple regression analysis and found that the regularity of learning intervals was an important predictor of final grade while total time and login frequency were not important predictors.

Lee et al. [23] used multinomial logistic regression to predict the final grade based on used features. Results showed that assignments are the strongest predictor of the final grade in face-to-face courses, and quizzes are the strongest predictor of the final grade in online courses.

This review found that all prediction studies are similar in predicting the final grade but differ in the type of predictor variable. For example, studies [19], [21] and [23] found that the assignment is the strongest predictor of the final grade whereas the strongest predictor in [20] and [22] are the type of question asked in the forum and regularity of learning intervals respectively. Regarding the methodology used, most of the prediction studies used logistic regression [19], [20], [21], [23] and multiple regression [19], [21] because of ordinal nature of predicted variable (the final grade).

3.2 Clustering
Clustering is a technique used to group a dataset into smaller subsets of similar objects or clusters [24]. Clustering can be

used to group learners depending on their behavior or learning difficulties. Different objects can be clustered in e-learning, such as students, courses, and content.

Romero et al. [24] grouped students based on their activities in the LMS. These activities include the number of assignments and quizzes and participation in discussions. Clustering was performed using the k-means algorithm to group students into three clusters: very active students, active students, and non-active students. Their results can be helpful for student grouping in collaborative activities.

Abdous, He and Yen in [20], categorized question themes using a hierarchical cluster analysis into four categories: class check-in, deadline/schedule, evaluation/technical and learning/comprehension.

Valsamidis et al. [25] categorized courses based on usage rate using the k-means algorithm. They found two clusters: high activity courses and low activity courses. Researchers proposed a way for measuring the quality of the course depending on the number of viewed pages by students for each session. Also, they investigated the relationship between the quality of the course and total course score.

Lust et al. [26] used k-means to group students based on their use of LMS and self-reported self-regulation strategies. They found four clusters reflecting tool used and strategy. Clusters are self-regulated and deeply oriented students, disorganized students, undefined students and inconsistent students.

In [27], Yildiz et al. cluster students depending on usage data (such as quiz score, exam score) using three clustering methods to find the best approach for estimating the final grade: k-means, fuzzy c-means, and subtractive clustering. They found that fuzzy c-mean is the best method for predicting student performance.

Lee et al. [23] applied expectation maximization and k-means to cluster courses depending on instructor and students usage data. Clusters differ in terms of assignments, quizzes, discussion topics and wiki pages. They found three students clusters and three instructor clusters based on LMS features used.

The results of clustering studies showed that the most commonly used algorithm was the k-means algorithm because it is one of the most popular and simplest algorithms used in data mining. Regarding the clustering object, students is the most used clustering object [24][25][26][27]. Another clustering objects were courses [25] and course contents [20].

3.3 Classification
A classifier is a mapping from a feature space X to a discrete set of labels Y [28]. Classification provides a collection of labeled patterns and tends to predict class label for new unlabeled patterns.

Chen et al. [29] used decision tree based on student behavior to discover students group with same characteristics and same reaction to a specific strategy. The decision tree will help the instructor to analyze student historical data weblog and to estimate the effectiveness of a certain strategy to achieve the desired pedagogical goals.

Minaei-Bidgoli and Punch [30] proposed an approach to classifying students based on extracted data from educational system log to predict their final grade. They used a combination of four classifiers to increase the accuracy of the classification. Results show that the total number of correct answers is the most important features of the classification.

Kotsiantis et al. [31] used Naïve Bayes to detect learner's performance based on demographic information and written assignments. This will help the instructor to construct the software support tool for enhancing students' performance.

Cocea and Weibelzahl [32] investigate if the log data can estimate the motivational level of the student. Using the decision tree, the researchers find that the reading time was an important factor to detect the motivation.

Hamalainen and Vinni [33] used classification to detect student success in the middle of the course. They used Naïve Bayes based on exercise points to detect the performance as early as possible.

Romero et al. [24] used C4.5 algorithm to classify students into different groups according to final mark based on their activities in the LMS. The results can be used to make decisions to improve course activities or to detect learners with learning problems.

In a summary, the review found that all classification studies are oriented toward the student's behavior. For example, studies [30],[31] and [33] classified students according to their learning performance whereas [24],[29] and [32] classified students based on their characteristics.

Regarding the algorithm used, most of the classification studies used decision tree [29], [30], [32], [24] and Naïve Bayes [31][33]. A decision tree is the easiest method in interpreting and explaining results and it needs a little effort for data preparation. Naïve Bayes is simple, fast and easy to implement.

3.4 Association Rule Mining
Association rule mining discovers the relationship between variables in databases. It identifies rules that discovered in databases based on some measures of interesting [34].

Zaiane [35] used association rule mining to build an agent that can recommend activities on the learning system. Based on the history of learner navigation, the material of the course will be improved. Recommendations will help the learner to navigate the materials by facilitate finding relevant resources and assist the learner to choose activities that improve their performance.

Lu [36] propose a framework of a recommender system to help students to find the wanted learning materials. Using association rule mining, this recommender system can help the students to learn more effectively by providing personalized learning materials.

Ramli [37] used association rule mining to optimize the course content. By applying Apriori algorithm to log data, they mine the relationship between visited pages. This will be helpful for an instructor to increase the performance of web services through improving the website in terms of its contents, structure, presentation, and delivery.

Markellos et al. [38] analyzed the weblog data using association rules to produce recommendations to users in the learning system. They used Apriori algorithm on learning materials to determine the most suitable one for the student.

Romero et al. [24] used the Apriori algorithm for association rule mining. Based on student's data from weblog, authors found some interesting rules that may help the instructor to pay attention to students that prone to failure and help them to success.

All studies used association rule mining are used for

optimizing the e-learning contents according to the students interesting. Contents such as activities [35], learning materials [36][38], web pages [37], messages [24] will recommend to the student for providing a personalized learning environment.

3.5 Distillation of Data for Human Judgment

Distillation of data for human judgment uses techniques for data representation to enable humans to understand it quickly and easily [18]. The most popular and effective technique is a visualization. Information visualization [39] is a computer graphics and user interface branch concerned with the representation of an interactive image to the user. This will facilitate the analysis of vast amount of information through charts, scatter plots or 3d representation.

Shen et al. [40] present a Data Analysis Center system to present student's data such as solved assignment and quizzes marks. This system can show the place of the learner in the class and define the weakness of the learning process. It can help teachers to analyze learner's patterns and to construct the materials efficiently.

Romero et al. [24] used GISMO tool to visualize data from Moodle log such as graphs representing. Course access, resources access, discussion participating. This will allow the instructor to be aware of what is happening in the learning process.

Valsamidis et al. [25] used a Markov clustering algorithm which combines cluster analysis and graphical representation to represent the relationship between students. The similar students were grouped visually to help interpreting results.

Arnoled and Pistilli [41] developed a Course Signals application that used a dashboard to provide a real-time display of student performance to instructors and students. They get a positive result on student's grades when using this application.

Thakur et al. in [21] examine LMS usage data to investigate the constancy in student's marks over the course by using heat maps. The result of the investigation was that the senior student's grades were stable whereas the freshmen grades tended to be unstable.

Lee et al [23] used clustergram of the features used by students in LMS to help visualization of student's behavior among the course and relate it to the final grade.

As a result, all studies of visualization are focused on students' performance in the learning process to get the better understanding of student's behavior and detect student which are at-risk.

Studies on EDM and their applications are listed in Table 1.

4. ADAPTIVE E-LEARNING AND GAMIFICATION TECHNIQUES

The ability to identifying students who are at risk of failing a course is important for teachers to take corrective actions [42]. An emerging type of learning is the adaptive e-learning. Each student has their preferences, current skill and learning styles [43]. Adaptive E-learning is designed to offer personalized experiences to students according to his personality. Gathering data, analyzing and interpreting skill level and preferences of the user and applying the necessary adjustments to the system in order to fit the current ability of the learner is the main feature of this approach.

Literature review modeled the adaptive learning environment from several perspectives [44]:

a) Domain model: during a learning process, this model focus on relationships between course material to adapt the content delivery and timing of courses.

b) Learner model: in this model, the system continues dynamically adjusted with respond to the user, his interaction with the system, his responses, and his personal characteristics.

c) Group model: this model analyzes groups of learners in individually find learning patterns in this groups.

d) Adaptive model: This model joins the adaptive theory of an adaptive learning environments ALE, at various levels of abstraction.

The concept of gamification is defined as "the incorporation of game design aspects in the implementation of software systems unrelated to games" [44]. Offering rewards, instant feedback, competition and freedom to fail to learners bring an additional layer of user engagement in the learning process.

Nowadays research is done to find out the probable relations between gamification-elements and learners' motivation [45], [46].

5. ROADMAP FOR USING DATA MINING AND GAMIFICATION IN ADAPTIVE E-LEARNING

Adaptive e-learning is an emerging type of learning. Powerful adaptive e-learning system needs to be well designed.

Some guidelines were proposed for designing an effective adaptive e-learning system:

First, using educational data mining EDM in adaptive e-learning systems will enhance the performance of E-learning system.

Second, finding out the probable relations between gamification elements and learners' motivation in Adaptive E-learning system is very important to improve the Learning system.

6. CONCLUSION

EDM and gamification are novel domains. They apply the methods and techniques of data mining and gamification in adaptive educational context. The goal of using EDM is to facilitate the decision making about the learning process to increase its effectiveness.

This paper surveys the state of the art of applications of data mining in e-learning. There are many studies conducted using different methods of data mining. Many studies used EDM for several goals such as predicting student's performances, grouping students according to their data, classify contents of the course, producing a recommendation for students and visualize student information.

Most of EDM studies results have to be interpreted effectively to help instructors in decision making. Decisions that can be made to increase the effectiveness of e-learning are such as adapting the course contents based on the student's categories, improving the course design according to students usage, giving feedback to the student to lead them to change their behavior, detecting students at-risk early depending on their assignment or quizzes marks.

Because of the fact that EDM is fast growing and its application was useful, It is expected that many applications will be proposed and implemented in the coming few years to help the learning process.

7. REFERENCES

[1] R. C. Clark and R. E. Mayer, "e-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning", 3rd ed. Pfeiffer, San Francisco, 2011.

[2] J. Mostow and J. Beck, "Some useful tactics to modify, map and mine data from intelligent tutors", Natural Language Engineering, vol. 12, no. 2, pp.195–208, 2006.

[3] O. Zaiaane and J. Luo, "Web usage mining for a better web-based learning environment", In Proceedings of the conference on advanced technology for education, Banff, Alberta, pp. 60–64, 2001.

[4] W. Klosgen and J. Zytkow, "Handbook of data mining and knowledge discovery", New York: Oxford University Press, 2002.

[5] J. Lee and J. Hammer, "Gamification in Education: What, How, Why Bother?", Academic Exchange Quarterly, 15(2), 2011.

[6] D. Ribeiro, M. Ventura, "What about a gamified learning environment?- Thinking on gamification challenges", Electronic International Interdisciplinary Conference, 2013, p. 568.

[7] V. Smith, A. Lange and D. Huston, "Predictive Modeling to Forecast Student Outcomes and Drive Effective Interventions in Online Community College Courses", Journal of Asynchronous Learning Networks, vol. 16, no. 3, 2012.

[8] T. Browne, M. Jenkins, and R. Walker, "A longitudinal perspective regarding the use of VLEs by higher education institutions in the United Kingdom", Interactive Learning Environments, vol. 14, no. 2, pp. 177–192, 2006.

[9] S. Dawson, "A study of the relationship between student communication interaction and sense of community", Educational Technology & Society, vol. 11, no. 3, pp. 224–238, 2008.

[10] T. Browne, M. Jenkins, and R. Walker, "A longitudinal perspective regarding the use of VLEs by higher education institutions in the United Kingdom", Interactive Learning Environments, vol. 14, no. 2, pp. 177–192, 2006.

[11] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601-618, 2010.

[12] Y. A. AlMazroui, "A survey of Data mining in the context of E-learning", International Journal of Information Technology & Computer Science, Vol. 7, no. 3, 2013.

[13] K. Koedinger, S. D'Mello, E. McLaughlin, Z. Pardos and C. Rosé, "Data mining and education", Wiley Interdisciplinary Reviews: Cognitive Science, vol. 6, no. 4, pp. 333-353, 2015.

[14] M. Recker and J. E. Lee, "Analyzing Learner and Instructor Interactions within Learning Management Systems: Approaches and Examples", Learning, Design, and Technology, 2016.

[15] Krumm, R. Waddington, S. Teasley and S. Lonn, "A Learning Management System-Based Early Warning System for Academic Advising in Undergraduate Engineering", Learning Analytics, pp. 103-119, 2014.

[16] G. Siemens, P. Long, "Penetrating the fog: Analytics in learning and education:, Educause Review, vol. 46, no. 5, pp. 30-32, 2011.

[17] M. Berland, R. Baker, and P. Blikstein, "Educational Data Mining and Learning Analytics: Applications to Constructionist Research", Technology, Knowledge and Learning, vol. 19, no. 1-2, pp. 205-220, 2014

[18] M. Bienkowski, M. Feng, and B. Means, "Enhancing teaching and learning through educational data mining and learning analytics: An issue brief". Washington, DC: US Department of Education, 2012.

[19] L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an "early warning system" for educators: A proof of concept", Computers & Education, vol. 54, no. 2, pp. 588–599, 2010.

[20] M. Abdous, W. He, and C. J. Yen, "Using data mining for predicting relationships between online question theme and final grade", Journal of Educational Technology and Society, vol. 15, no. 3, pp. 77–88, 2012.

[21] G. S. Thakur, M. M. Olama, A. W. McNair, S. R. Sukumar, and S. Studham, "Towards adaptive educational assessments: Predicting student performance using temporal stability and data analytics in learning management systems", In Proceedings 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York, NY: ACM, 2014.

[22] I.H. Jo, D. Kim, and M. Yoon, "Constructing proxy variables to measure adult learners' time management strategies in LMS". Educational Technology and Society, vol. 18, no. 3, pp.214–225, 2015.

[23] J. E. Lee, M. Recker, H. Choi, W.J. Hong, N.J. Kim, K. Lee, M. Lefler, J. Louviere and A.Walker, "Applying data mining methods to understand user interactions within learning management systems: Approaches and lessons learned", Journal of Educational Technology Development and Exchange, vol. 8, no. 2, pp. 99-116, 2016.

[24] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial", Computers & Education, vol. 51, no. 1, pp. 368–384, 2008.

[25] S. Valsamidis, S. Kontogiannis, I. Kazanidis, T. Theodosiou, and A. Karakos, "A clustering methodology of weblog data for learning management systems", Journal of Educational Technology and Society, vol. 15, no. 2, pp. 154–167, 2012.

[26] G. Lust, J. Elen, and G. Clarebout, "Regulation of tool-use within a blended course: Student differences and performance effects", Computers & Education, vol. 60, no. 1, pp. 385–395, 2013.

[27] O. Yildiz, A. Bal, and S. Gulsecen, "Statistical and clustering based rules extraction approach for a fuzzy model to estimate academic performance in distance education", Eurasia Journal of Mathematics, Science and Technology Education, vol. 11, no. 2, pp. 391–404, 2015.

[28] R. O. Duda, P. E. Hart and D.G. Stork, "Pattern classification". Wiley Interscience, 2000

[29] G. Chen, C. Liu, K. Ou and B. Liu, "Discovering decision knowledge from weblog portfolio for managing classroom processes by applying decision tree and data cube technology", Journal of Educational Computing Research, vol. 23, no. 3, pp. 305–332, 2000.

[30] B. Minaei-Bidgoli and W. Punch, "Using genetic algorithms for data mining optimization in an educational web-based system", In Genetic and evolutionary computation conference, Chicago, USA, pp. 2252–2263, 2003.

[31] S. B. Kotsiantis, C. J. Pierrakeas and P. E. Pintelas, "Predicting students' performance in distance learning using machine learning techniques", Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, 2004.

[32] M. Cocea and S. Weibelzahl, "Can log files analysis estimate learners' level of motivation?", In Proceedings of the workshop week Lernen - Wissensentdeckung - Adaptivitate t, Hildesheim, pp. 32–35, 2006.

[33] W. Hamalainen and M. Vinni, "Comparison of machine learning methods for intelligent tutoring systems", In Proceedings of the eighth international conference in intelligent tutoring systems, Taiwan, pp. 525–534, 2006.

[34] R. Agarwal, T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases", In Proceedings of the ACM SIGMOD international conference on management of data, Washington DC, USA, pp. 1–22, 1993.

[35] O. Zaiane, "Building a recommender agent for e-learning systems", In Proceedings of the international conference on education, Auckland, New Zealand, pp. 55–59, 2002.

[36] J. Lu, "Personalized e-learning material recommender system". In International conference on information technology for applications, Utah, USA, pp. 374–379, 2004.

[37] A. A. Ramli, "Web usage mining using apriori algorithm: UUM learning care portal case", In International conference on knowledge management, Malaysia, pp. 1–19, 2005.

[38] P. Markellou, I. Mousourouli, S. Spiros, and A. Tsakalidis, "Using semantic web mining technologies for personalized e-learning experiences". In Proceedings of the web-based education, Grindelwald, Switzerland, pp. 461–826, 2005.

[39] R. Spence, "Information Visualization", Addison Weslet, 2001.

[40] R. Shen, F. Yang, and P. Han, "Data analysis center based on e-learning platform", In Workshop the internet challenge: Technology and Applications, Berlin, Germany, pp. 19–28, 2002.

[41] K.E. Arnold, and M.D. Pistilli, "Course signals at Purdue: Using learning analytics to increase student success", In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 267–270, Vancouver, BC: ACM, 2012.

[42] W. Han, D. Jun, G. Xiaopeng, L Kangxu, "Supporting Quality Teaching using Educational Data Mining based on OpenEdX Platform"2017 IEEE Frontiers in Education Conference (FIE)Year: 2017Pages: 1 – 7.

[43] Elena Madalina Jianu, Andrei Vasilateanu, "Designing of an e-learning system using adaptivity and gamification"2017 IEEE International Systems Engineering Symposium (ISSE), Pages: 1 – 4.

[44] A. Paramythis and S. Loidl-Reisinger, "Adaptive Learning Environments and e-Learning Standards", ECEL 2003, ed. 2, 2003.

[45] Fatemeh Roosta, Fattaneh Taghiyareh, Maedeh Mosharraf, "Personalization of gamification-elements in an e-learning environment based on learners' motivation ", 2016 8th International Symposium on Telecommunications (IST), Pages: 637 – 642.

[46] L. S. Ferro, S. P. Walz, and S. Greuter, ―Towards personalized, gamified systems : an investigation into gamification design , personality and player typologies", IE '13 Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death, Melbourne, Australia ― September 30 - October 01, 2013, pp. 1–6, 2013.

8. APPENDIX

Table 1. EDM methods and applications

Data Mining Method	Paper	Year	Application
Prediction	[19] Macfadyen and Dawson	2010	Predicting the final grade based on posted topics in forum, sent mail and completed assignment
	[20] Abdous et al.	2012	Predicting the final grade based on online question theme
	[21] Thakur et al.	2014	Predicting the final grade based on number of completed assignment
	[22] Jo et al.	2015	Predicting the final grade based on regularity of learning intervals
	[23] Lee et al.	2016	Predicting the final grade based on assignment participation in face-to-face courses and on quiz participation in online course
Clustering	[24] Romero et al.	2008	Clustering students into three clusters: very active students, active students, and non-active students
	[20] Abdous et al.	2012	Clustering question theme into four clusters: class check-in, deadline/schedule, evaluation/technical and learning/comprehension
	[25] Valsamidis et al.	2012	Clustering course into two courses clusters: high activity courses and low activity courses.
	[26] Lust et al.	2013	Clustering students into four clusters: self-regulated and deeply oriented students, disorganized students, undefined students and inconsistent students.
	[27] Yildiz al.	2015	Clustering students into eight clusters using k-means, nine clusters using fuzzy c-means and eleven clusters using subtractive clustering
	[23] Lee et al.	2016	Clustering instructors into three instructor clusters Clustering students into three student clusters
Classification	[29] Chen et al.	2000	Discovering students group with same characteristics and same reaction to a specific strategy to estimate the suitable strategy.
	[30] Minaei-Bidgoli and Punch	2003	Predicting student performance and final grade
	[31] Kotsiantis et al.	2004	Detecting learners performance and determining the support tool
	[32] Cocea & Weibelzahl	2006	Identifying learners motivation based on students behavior to make an intervention
	[33] Hamalainen and Vinni	2006	Detecting the performance of learner as early as possible
	[24] Romero et al.	2008	Estimating the suitable course activities based on learner behavior
Association rule mining	[35] Zaiane	2002	Recommending activities on the learning system.
	[36] Lu	2004	Recommender system providing personalized learning materials.
	[37] Ramli	2005	Optimizing the course content for instructor to increase the performance of web services
	[38] Markellou et al.	2005	Producing recommendation to users in learning system
	[24] Romero et al.	2008	Paying attention to students that prone to failure
Distillation of Data for human judgment	[40] Shen et al.	2000	Analyzing learners patterns and constructing the materials efficiently.
	[24] Romero et al.	2008	Helping instructor to be aware of what is happening in the learning process.
	[25] Valsamidis et al.	2012	Represent the relationship between students.
	[41] Arnoled and Pistilli	2012	Providing a real-time display of student performance to instructors and students
	[21] Thakur et al.	2014	Investigating the constancy in students marks over the course
	[23] Lee et al	2016	Visualizing of students behavior among the course