Quarterly of Iranian Distance Education Journal
Vol.2, No.1, Serial Number 5, Summer 2019 (P 33-42), Payame Noor University

Exploring Learners' Behaviors in Interactive Educational Environments by Data Mining Algorithms

Maryam Haghshenas[1*], Fatemeh Torabi Konjin [2], Alireza Namjoo[3]
1. Ph.D. of Media Management of, Tehran University
2. M.A. Information Technology Engineering
Islamic Azad Tehran University
3. M.A. Student of Information Technology Engineering
Tarbiat Modarres University

Received: 2019/4/29 **Accepted:** 2019/6/30

Abstract
Data mining as a science is a search process for finding new, valuable, and unknown information among a wealth of data. This approach is among the top 10 developing knowledge that will face the next decade with a technological revolution and has been expanding rapidly in recent years. The purpose of this research is to explore learners' behavior in interactive educational environments using data mining algorithms. Exploration and extraction of meaningful relationships among different factors with the help of data mining algorithms can be used to provide solutions to improve learners' performance. The research instrument was a questionnaire. The validity of the questionnaire was obtained through content analysis and its reliability was measured through re-testing. The data collection discussed in this study relates to the activities and interactions of 200 active students in a university social network. After gathering information about learners' interactions, the Apriori algorithm of data mining is used to analyze the behavior of learners through the use of rules. The main goal is to facilitate the education along with obtaining the proficiency of the courses. Improved interactive content is obtained from feedback information based on the data mining techniques of the data management system. Therefore, the results of this research can help educational system managers to plan their education and optimize educational processes in the interactive environment.

Keywords
Behavioral Sciences, Interactive Learners, Data Mining Algorithms, Association Rules.

Introduction
The unparalleled emergence and growth of the Internet and the web has made its applications to have a great positive impact on everyday human life. One of these areas is e-learning. This kind of education has made it possible for all people of the community to study regardless of time and place positions. On the other hand, data mining in education can have a great impact on the quality of education and the use of its techniques can improve the ability to use the conditions of education [4].

Data mining or the science of extracting knowledge from the database addresses the discovery of innovative and useful patterns from among a massive amount of information available in the mass of data stored in the database [8], and using powerful approaches such as the science of statistics, data patterns recognition algorithms, neural networks and various machine learning algorithms, it is capable of discovering valid knowledge from a wealth of data, thus ultimately producing decision-making rules [10].

Nowadays, given the high volume of stored data pertaining to students in the existing databases of education Department and the increasing upgrading of these data, obtaining the

*Corresponding Author: M_haghshenas@ut.ac.ir

knowledge contained in these data has become a necessity for improving the quality and proper planning of education that we use data mining techniques to solve this problem [10].

Data mining techniques have been expanded so that they can easily be applied to modern software tools, and the information gathered can be utilized best. With the help of these techniques and data mining algorithms, an active electronic learning pattern can be achieved with effective modeling [4].

This process in an educational environment, with the goal of discovering hidden educational patterns for improving the performance of the educational system, can be applied and is widely used in various fields, including education. Universities that set up social networks or interactive environments such as blogs, forums, and the like aiming at learning and educating, produce massive amounts of information by web servers, and gather it in the area of events section. Web-based environments provide the ability to record and store information from students' learning behaviors in their personal profiles. Analyzing and examining the data of learners' interaction with the web-based learning environment is done by data mining [4].

In this research, using the Apriori algorithm, an interactive learner's behavior is analyzed and discussed. This article is organized as follows. In Sections 2 and 3, after describing the data mining technique, the literature reviewed in this area is discussed; in Section 4 of the Apriori algorithm and the association rules will be described in detail. In Section 5 of this research, the methodology of the research is fully explained, in Section 6, the findings of the research are presented based on the data mining algorithm and finally, in Section 7, the results of this study are expressed explicitly.

Educational Data Mining

Specifically, in the field of education, data mining techniques have been used to examine key questions and to find empirical evidence for the development of theoretical education, which has led to the emergence of a new field of research called educational data mining. Educational data mining is used to address various issues related to education and learning [8]. Educational data mining is a branch of data mining that provides data mining tools and techniques for using data in the education sector and helps researchers to better understand the motivations and interests of learners in the learning process [1].

Using data mining in education is a repetitive cycle. Each educational system stores a large amount of data based on the history of user interactions with systems. These data are a good source of knowledge. Using data mining techniques, we can analyze the data and create appropriate data sets and patterns. By adding more features to these models, we can achieve more success in the effective learning process, including: personalizing the learning process, providing feedback to producers of educational content, creating student groups, and identifying degree of fraud among each other. Educational data mining is more complex than data mining in any field (for example, in business). Students may learn a lesson several times in a course, visit different sites and other sections in the course, play games and work in long periods of time. Receiving and defining the inputs to the student system is one of the biggest challenges [4].

Finding the patterns and knowledge that lies behind the various factors in students in the educational environment can help educational decision-makers to improve and promote educational processes such as planning, registration, assessment and counseling, and thus, helping managerial personnel plan better and have more accurate measures to guide the students. The extracted knowledge will be useful not only for educators and managers, but also for the learners themselves. To have an advanced plan and a successful planning requires comprehensive awareness and careful decision making and goal setting. The scholars of psychology believe that man develops at moments that he decides. An important condition for growth and development is familiarity with the skills and factors surrounding the environment for decision making [10].

Literature Review

In recent years, in order to optimally use the large amount of information stored in educational systems and to extract knowledge from this information, there have been plenty of activities in educational systems that can be solved through data mining. The following can be mentioned which are based on data mining techniques in this area [11] :

- In Romero and Ventura's study in 2007, a comprehensive review of the research on this area from 1995 to 2005 was presented, and a summary of the achievements of each research is presented in the study.
- In Minaea et al research (2012), the identification of factors affecting academic failure was studied using association rules and cluster analysis. In this research, attempts have been made to implement predictive data mining models in order to predict students' academic status based on their personal characteristics and their educational past.
- In a study by Bhardwaj and Pal (2012), it has been verified that students' academic performance is based on various factors such as individual, social, psychological, and other environmental factors, and the results suggest that by using the Bayes' classification which is simple, it is believed that many factors have a significant impact on the performance of students, and identification of these factors can affect their performance in the future.
- According to Luan (2002), using two-step clustering algorithms, Decision Tree Algorithms and Neural Networks on the data of 1,500 students, researchers were able to cluster and predict enduring and non-enduring students.
- According to Quadri and Kalyankar in 2012, evaluating and examining various factors such as students' gender characteristics, presence in the classroom, parents' income, being the first child, etc. have been dealt with in dropping educational institutes by using the Decision Tree .
- In the study of Kotsiantis and his colleagues in 2004, the efficiency of students in the educational environment, along with the relevance of the influencing characteristics in improving it, was studied.
- Hamalainen and Vinni used association rules to address the problems of students in the educational setting and provide counseling to them.
- In Zaiane's research, association rules were used to guide students' activities and offer educational content. In a study by Vialardi et al. (2011), data mining techniques were used with the CRISP methodology to design a system of guidance in support of the student registration process in accordance with the students' educational backgrounds.
- A Bayesian Network Model has been proposed by Torabi et al (2012) to predict students' grades, taking into account students' characteristics and their educational backgrounds.
- In their research, Skooea and Asgari (2014), used data mining techniques to explore knowledge by collecting information from students' behavior in order to find students' behavioral relationships and academic performance.
- Baker and Yacef (2009), have identified students' inappropriate use of the learning environment and their irresponsible behavior using class rating techniques [10] .
- In 2006, Tahir and Naqvy did studies over the performance of students and concluded that the proportion of attendance in the classroom, hours of study on a daily basis after university, income, family, mother's age, and the level of literacy of students' mothers were related to the degree of student performance [10].
- In their study, Attaway and Bry (2004), examined the relationship between students' academic achievement and their parents' income, and found that parents' incomes had a positive impact on student progress.
- In Shannaq et al research (2010), using data mining, the factors influencing student enrollment in universities were studied and the characteristics of loyal students were identified. They studied the law-producing process based on the decision-making process as a rating method and by using different methods, and identified the main characteristics that may affect students'

loyalty [11] .

- Maghsoudi et al (2012) investigated students' behaviors and predicted scores at the E-Learning Center of Iran University of Science and Industry by using association rules and the Decision Tree. The results were highly accurate and the aim was to use these experiences in an e-learning system to guide academic performance [8].
- Dey pir and Rabo (2018) studied the use of clustering approach and k-means algorithm to classify learners in e-learning environments in order to provide specific educational styles. The proposed model was evaluated for better educational planning in real time, and the results showed that the proposed model has high efficiency. [3].
- Haiyan Jia et al (2011), using ontology and Semantic Web, provide a mechanism for interpreting educational content, reusability of educational content, combining educational resources and qualifying content of instruction. The results indicate that the use of an educational system with high quality content along with an ontological approach will increase learning by 20% and increase the interest in using the educational system by 40% [5].
- Baghaei et al (2017), using two decision tree techniques and neural networks on student data from a university in Mashhad, investigated educational violations and the impact of such variables as gender, field of study, etc., with the type of violation. The results show that except for gender, other variables, such as the field of study, the educational level, the average of the previous half-year, the number of units and the term of the violation affect the type of violation [1].
- Jahed Saroni et al (2016), utilizing predictive Decision Tree Model, Closest Neighbors, Naive Bayesian, and Neural Network, identified unapt or slow-learning student. The decision tree model in this connection has had the highest accuracy of 95.92% and can be used in schools and educational institutions in the prediction programs of the students. The validation data collection is being compiled and analyzed from the real world of art schools using the Rapid Miner software [7].
- Tabrizi et al (2017), predicted some of the factors affecting academic failure by accessing data from Tehran's students' college students and data mining algorithms. First, the k-means clustering method and then a few methods of the classification classes are implemented and the results of the classification algorithms are compared. The final decision tree, relative to other trees, is based on the optimality and accuracy criterion. The present study is applied in terms of purpose. The results of this study will help the education managers to prevent students from dropping out based on the patterns they have discovered [9].

Association Rules and Apriori Algorithm

Association rules are commonly used in data mining in exploring repetitive patterns, identifying affiliations, and the relationship between variables in a data set. The general form of these rules is expressed as A, then B, which expresses the simultaneous occurrence between A and B, which are referred to as the antecedent and consequent of the rules. Different criteria are used to measure the validity and worthiness of the rules whose most widely used are two criteria of Min-Support and Min-Confidence, and rules that can meet the min-support and the min-confidence are strong rules.

Relationships (1) and (2) express the association rules with two criteria of support and confidence.

$(A \cap B = \emptyset)\ A \Rightarrow B$

$$\text{Support}\ (A \Rightarrow B) = \frac{|A \cup B|}{N} \qquad (1)$$

$$\text{Confidence } (A \Rightarrow B) = \frac{|A \cup B|}{|A|} \qquad (2)$$

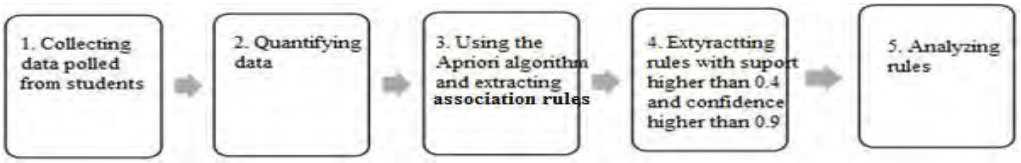
The degree of support represents the total number of transactions in the entire data set, which includes both the values of A and B, and the degree of confidence indicates the simultaneous frequency of repetitions of A and B with the frequency of repetitions of A alone. The Apriori algorithm, FP-Growth algorithm and genetic algorithm are algorithms for extracting association rules. The Apriori algorithm is the most used algorithm proposed by Agrawal and Srikant in 1994, which explores the collection of items. This algorithm uses a repeatable method called Level-wise, which makes use of a set of n-member items to explore the collection of n + 1 member items. In this way, firstly, the set of single-member items is found by scanning the database, and those items that meet the min-support and the min-confidence degrees are maintained. Then, using them, frequent two-member sets of repetitive items, three- member and n-member sets are extracted. An important advantage of this algorithm versus other algorithms for generating repetitive items is its simplicity of implementation, and its weakness is the need for more memory to hold input transactions as well as data rotation each time the algorithm is run. This algorithm has an anti-monotony property, that is, any set of items that fails to meet the min-support and the min-confidence and is not repetitive, will have all its subsets not be replicated [6].

Research Method

Since the purpose of the research is to study the behavior of learners in interactive educational environments, and due to the increasing use of social media in academic environments and educational centers, social networks are recognized as one of the interactive tools in learning environments. The approach of this research is descriptive and to collect data on the level of engagement and participation of students in social networks, we attempted to come up with a data collection questionnaire. In preparing the 20-question questionnaire, a five-point Likert Scale was used with very low, low, moderate, high and very high points. The validity of the data collected from the questionnaires was confirmed by the education and training specialists and the university administrators. The questions and answers given by students are shown in Table 1. In general, the purpose of the present research is to investigate the interactive learner behavior in interactive learning environments in the form of association rules by using the Apriori algorithm. The statistical population of the study consists of 200 students active in the academic Social network. In addition to attending classrooms, in this interactive environment they also share booklets and educational files with teachers and classmates and discuss and exchange and submit comments and create educational interactions.

This research is summarized in the following 5 steps based on Fig. 1.

Figure1. Research Steps



Tabel1. Questions and Frequency of Answers to the Researcher-made Questionnaire

Questions	Frequency of answers				
	Very low	***Low***	***Moderate***	***High***	***Very high***
How much do you study per day?	15	74	74	25	12
How much do you use the Internet during the week?	0	1	95	91	13
How much do you work with your computer?	0	1	95	91	13
How much are you interested in the virtual world?	0	1	95	91	13
To what extent are you able to establish friendly interaction with other students?	0	0	53	73	74
To what extent are you engaged in curricular discussions?	0	0	0	124	76
How much are you interested in scientific discussions?	0	0	0	126	74
To what extent are you interested in the professor's manner of teaching?	6	90	94	0	0
What do you think of professors' dominance over curricular material?	15	74	74	25	12
How much public relations and ability do you have in establishing communication with others?	0	0	53	73	74
Are you willing to establish scientific and research communications with students and professors?	0	0	31	95	74
Are you willing to express your views in an interactive environment?	0	0	36	68	96
Are you interested in collaboration and co-thinking with other students?	0	0	32	71	97
To what extent are you interested in utilizing peoples' knowledge in an interactive environment?	0	0	26	68	106
To what extent are you interested in using professors' experiences?	0	0	26	68	106
Are you interested in using other students' ideas?	0	0	32	71	97
Are you interested in sharing your knowledge with other students?	0	0	26	88	86
Are you interested in using other students' educational experiences in a network?	0	0	31	75	94
How much do you understand the scientific characteristics of the classmates present in the network?	0	0	31	95	74
How do you evaluate the thinking movement by professors?	10	49	77	40	24

Research Findings

After collecting the data from the students' questionnaires, qualitative data were turnedinto quantitative data based on the Likert scale. Then, by implementing the Apriori algorithm, one of the most widely used algorithms for association rules technique, 140 laws were produced with a minimum number of 40% for the support criterion and a minimum number of 90% for confidence; in Table 2, a number of rules that are attractive more were selected and displayed.

Tabel2. Selected Rules with the Highest Attraction

Row	***Antecedent***	***Consequent***	***% Support***	***Confidence%***
1	Students who use the network and are very interested in collaborating with other students.	They are very interested in using other students' educational experiences in the network.	46	95
2	Students who are on the net and are very interested in using the ideas and thoughts of other students.	They are very interested in using other student experiences in the network.	45	95
3	Students who use the Internet on a moderate level a nd are in an interactive environment.	Their interest in virtual space is moderate.	47	100
4	Students who use the network and are very interested in using the experiences of professors in an interactive environment.	They are very interested in using the knowledge of individuals in an interactive environment.	50	100
5	Students who are highly interested in the network to communicate with other students and are well aware of their scientific qualities.	They are keenly interested in sharing their knowledge with others.	43	91
6	Students who are interested in networking and sharing their knowledge with other students	They are keenly interested in communicating with students.	42	98
7	Students who are keenly interested in expressing their views and opinions and are highly interested in engaging with other students.	They are very interested in using the ideas and thoughts of other students.	42	100
8	Students who use the Internet on a large scale during the week and are highly interested in the virtual space and using an interactive environment.	They use computers a lot during the day.	45	100

This section describes the rules extracted in Table 1. These rules are selected based on high numbers of support and confidence.

Law 1: Students who use the interactive environment and are very interested in engaging and collaborating with other students, then are very interested in using other students' educational experiences in the interactive environment (support = 46% and confidence = 95 %) This law suggests that when students use the interactive environment and are very interested in engaging with other students, they are likely to have an interest in using other students' educational experiences in an interactive environment for 95% and this law has a 46% support.

- Law 2: Students who use the interactive environment and are very interested in using the ideas and thoughts of other students, then are very interested in using the educational experiences of other students in the interactive environment (support = 45% and confidence = 95%). This law suggests that students who use the interactive environment and have a great interest in using the ideas and thoughts of other students are likely to have an interest in using the educational experiences of other students in an interactive environment for 95%, and this law has a 45% support.
- Law 3: Students who use the Internet on a moderate level and use the interactive environment, then their interest in using the interactive environment and presence in the cyberspace is moderate (support = 47% and confidence= 100%). The interpretation of this law suggests that students who use the interactive environment and generally use the Internet at a moderate level are 100% likely to have moderate interest in using the interactive environment and this law has a 47% support.
- Law 4: Students who use the interactive environment and are very interested in using the experiences of professors in an interactive environment, then are very interested in exploiting the knowledge of individuals in the interactive environment (support = 50% and confidence = 100%). This law suggests that students who use the interactive environment and have a great interest in using the experiences of professors in an interactive environment are likely to have a 100% interest in utilizing the knowledge of individuals in an interactive environment, and this law has a 50% support.
- Law 5: Students who use the interactive environment and are keenly interested in communicating with other students and are familiar with other scholars' academic qualitiesto a large extent, then are highly interested in sharing their knowledge with others (support = 43% and confidence = 91%). The interpretation of this law suggests that students who use interactive environments and are keen in scholarly communication with professors and other students and who know a great deal about the scientific characteristics of students share their knowledge with others enjoy a likelihood of 91% and this law has a 43% support.
- Law 6: Students who use the interactive environment and are keenly interested in sharing their knowledge with other students, then are greatly interested in communicating with students (support = 42% And reassurance = 98%). The law states that students who use the interactive environment and their level of interest in sharing their knowledge with other students is higher are 98% likely to be interested in establishing scientific and research communications with other students and this law has a 42% support.
- Law 7: Students who are very interested in expressing their views in an interactive environment, and also interested in engaging with other students, then are highly interested in using ideas and the thoughts of other students (support = 42% and confidence= 100%). This law suggests that 100 percent of students are more likely to be interested in using the ideas and thoughts of other students if their interest in expressing their views and point of view in the interactive environment is very high and they are keenly interested in collaborating with other students.
- Law 8: Students who use the Internet a lot during the week and are highly interested in being in cyberspace and using the interactive environment, then use to a large extent computer

throughout the day (support = 45% and confidence= 100%). This law suggests that, with a likelihood of 100%, students will use computers a lot during the day, if their Internet usage is high during the week and they are highly interested in cyberspace and also use the interactive environment.

Conclusion

Given the high volume of stored data relating to students in the existing databases in the education system and the increasing up-to-date of these data, obtaining the knowledge contained in these data is known a necessity for improving the quality and proper educational planning. Educational data mining as a science in the distance learning method is proposed, which is rich in data, methods and tools and is used in various educational environments for users. The analysis of learner behavior in social networks, which today is regarded as a new infrastructure for interactive educational environments, guides us towards a clever way of presenting learners' activities reporting, and leading the authorities of universities and educational and planning centers to use these environments in order to optimize learning processes and achieve educational goals is a more effective way. The use of data mining techniques in educational environments requires the efforts of IT professionals, educators and learners. The results of this research showed that university authorities should use interactive environments in order to increase knowledge of learners and to strive to build a specific culture for them.

References

[1] Seraji, Farhad; Attaran, Mohammad (2011). E-Learning: Basics, Design, Implementation and Evaluation. Hamedan: Bu Ali Sina University. [1] Z, Asghari and M, Hossein Nejad." Application of Data Mining Methods in Distance Education," CEAI Andisheh. The 3rd Regional Conference on Control, Electronics and Artificial intelligence. Iran, Tehran, February 2016.

[2] S, Baghaea and A.H, Ebrahim Rezaei Nik. "Investigation of students' educational misconduct by decision tree and neural network".3rd International Conference on Industrial Engineering and Management. Iran, Tehran, Ardabil University, January 2017.

[3] M, De Pir and A, Rabo." Using educational data mining to group learners into the e-learning environment to personalize the training program" .Journal of Management and Planning in Educational Systems.2018, pp.83-108.

[4] F, Ghasemi and A, Mirroshandel." Data mining in e-learning. International Conference on Applied Research in Information Technology, Computer and Telecommunications" International Conference on Information Technology, Computer & Communication, Iran, Torbat Heydarieh, Khorasan Razavi. November 2015.

[5] H, Jia, M,Wang and W,Ran." Design of a performance-oriented workplace e-learning system using ontology" Expert Systems with Applications, April 2011, pp3372-3382

[6] J, Han, M, Chamber and J, Pei." Data Mining Concepts and Techniques"2012, [3rd ed.USA: Waltham]

[7] A, Jahed Saroni and J, Hossein Khani." Identifying unapt students in the arts schools using the classification technique in data mining" 1sd National Conference on Advanced Technologies in Engineering. Iran, Birjand, Azad University, December 2016

[8] B, Maghsoudi, S, Soleimani, A, Amiri and M, Afsharchi." Improvement of education quality in e-learning systems using educational data" .Journal of Educational Technology, 2012.

[9] Z, Tabrizi, S, Jalili and R. Ehsani," A model for reducing educational dropout by data mining approach (Case study: Tehran Vocational Schools)"). International Congress on Improving Management and Education System, Iran, Tehran, Center for Datis Development Conferences, January2017

[10] M.R, Zahedi Fard, I, Attarzadeh and H, Pazokhzadeh." The study of high school students' behaviors with data mining techniques" International Conference on Applied Research in Information Technology, Computer and Telecommunications, Iran, Torbat Heydarieh, Khorasan Razavi Telecommunication Company. November 2015.

[11] M.R, Zahedi Fard, I, Attarzadeh and H, Pazokhzadeh.” Prediction of High School Students Performance by Data Mining Classification Techniques” International Conference on Applied Research in Information Technology, Computer and Telecommunications, Iran, Torbat Heydarieh, Khorasan Razavi Telecommnuication,November 2015.