Research Article

Minshun Zhang, Juntao Fan, Amit Sharma, and Ashima Kukkar*

Data mining applications in university information management system development

<https://doi.org/10.1515/jisys-2022-0006>

received October 08, 2021; accepted November 23, 2021

Abstract: Nowadays, the modern management is promoted to resolve the issue of unreliable information transmission and to provide work efficiency. The basic aim of the modern management is to be more effective in the role of the school to train talents and serve the society. This article focuses on the application of data mining (DM) in the development of information management system (IMS) in universities and colleges. DM provides powerful approaches for a variety of educational areas. Due to the large amount of student information that can be used to design valuable patterns relevant to student learning behavior, research in the field of education is continuously expanding. Educational data mining can be used by educational institutions to assess student performance, assisting the institution in recognizing the student's accomplishments. In DM, classification is a well-known technique that has been regularly used to determine student achievement. In this study, the process of DM and the application research of association rules is introduced in the development of IMS in universities and colleges. The results show that the curriculum covers the whole field and the minimum transaction support count be 2, $\min_{\text{conf}} = 70\%$. The results also suggested that students who choose one course also tend to choose the other course. The application of DM theory in university information will greatly upsurge the data analysis capability of administrators and improve the management level.

Keywords: data mining, universities and colleges, information management system

1 Introduction

With the advancement of the modern management theory and decision-making science, as well as their application in university and college management, universities and colleges will shift from experience-based management to scientific or information management based on the modern management theories and methodologies [1]. In light of the current situation, the information management method has been consistently supported in universities and colleges, and teaching information management system (IMS) has been established one after the other [2]. However, as the number of students managed and the time spent on using the teaching management information system increases, a large amount of management

* **Corresponding author: Ashima Kukkar**, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India, e-mail: ashima@chitkara.edu.in

Minshun Zhang: School of Economics and Management, Harbin University of Science and Technology, Heilongjiang, Harbin, 150040, China, e-mail: MinshunZhang@126.com

Juntao Fan: School of Economics and Management, Harbin University of Science and Technology, Heilongjiang, Harbin, 150040, China, e-mail: JuntaoFan1@163.com

Amit Sharma: Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India, e-mail: Sharma.amit@chitkara.edu.in

data based on teaching is collected. Education has become more flexible and broad as college enrollment continues to expand [3]. Most universities and colleges are dealing with a conflict between rising student numbers and tightening teaching resources, posing unprecedented difficulties to education management. As a result of such a novel occurrence, the IMS is growing in popularity and recognition among teachers and students because it allows them to work together effectively [4]. In the previous teaching management, school education management (SEM) only focused on the unique characteristics of the educational field, exaggerating the unique characteristics of school education, emphasizing the management mode based on experience, and to some extent ignoring the similarities between education and general management [5]. Further, SEM only emphasizes the particularity of education field and ignores the commonness between general and education management to some extent. It places too much emphasis on the precision of school curriculum and not enough on the experience-based management approach. For example, most universities and colleges have not established the IMS. It is an important modern management tool; but in sharp contrast, many universities and colleges themselves are developing various management information systems for the government and enterprises [6]. Many universities still do not realize the importance of management information system in university management [7]. Traditional data analysis approaches have some drawbacks in the face of “mountain” data collecting, whether in time or location, such as the inability to cope and unavailability of understandable administrators. There is a need to use these data effectively because it leads to increasingly more serious “data disaster,” forcing school administrators to take “decisions disaster” responses. In fact, whatever countermeasures are used, they are motivated by a sense of helplessness. Data mining (DM) technology advances the use of data from basic querying to more advanced applications such as prediction, decision assistance, and analysis [8]. The IMS has gradually been applied to university education as science and technology, as well as information technology, have continued to develop and improve. DM is a process of cooperation among various experts and also a process of high investment in capital and technology. This process should be repeated. In the process of repetition, the essence of things is constantly approached, and solutions to problems are constantly given priority [9]. The subdivision and reorganization of data process add and split the selected records. These data records are chosen from the data exploration clustering by analyzing the neural network (NN), decision tree (DT) mathematical statistics, and time series visualization. Further, a comprehensive interpretation is found to be effective in evaluating data knowledge, data sampling, data exploration, and data adjustment modeling [10]. Most universities and colleges now have an IMS in place, which have essentially overcome the problems and drawbacks of an old-fashioned teaching management [11].

1.1 Motivation

In universities and colleges, the purpose of teaching information management (TIM) is to maximize the utilization of teaching resources and leverage the benefits of diverse resources to attract high-skilled, high-quality, individuals for the country and society [12]. DM technology is a novel study area with a wide range of applications produced by the integration and intersection of multiple fields. It integrates database methods, artificial intelligence (AI) approaches, mathematical statistics approaches, and visualization approaches. With the expansion of the educational organization and the continuing information technology development, higher education is receiving an influx of teaching resources, and rigorous standards for higher education teaching management are also being imposed [13]. How to better and more effectively use information technology to service TIM and decision making has become a key indicator of a school strength and college or university’s teaching management level. The popularization and the use of DM technology is a research into the informationization of teaching management at universities and colleges in the context of a new circumstance that has significant real implications [14]. Implementing scientific DM methods and practices will undoubtedly become the aim and primary job of the future stage of education management informationization and virtual campus development in Chinese institutions [15].

In view of this research problem, Mago and Giabbanelli believe that DM is a way to extract hidden and possibly usable information and knowledge from a vast amount of incomplete, noisy, fuzzy, and random

data [16]. Li et al., conducted extensive study into the use of fuzzy approaches in the finding of knowledge [17]. Yanhao et al., studied data cube algebra [18]. On the basis of current research, this article introduces DM and its process, studies the application of association rules in the development of university information management system (UIMS), and shows that the application of DM theory in university information will greatly increase the data analysis ability of managers and improve the management level [19].

The following is a breakdown of the structure of this research article. Section 2 illustrates the related studies on educational DM. Section 3 describes the methods of DM in UIMS. Section 4 contains the technological process of UIMS based on DM technology. Sections 5 and 6 explore results, experimental test, and conclusion.

2 Related studies

DM technique is the most accurate method for evaluating usable data in the data warehouse [21]. DM is utilized to forecast hidden information using an extraction process in terms of improving decision making [22]. The use of DM for educational activities has been increased based on staff decisions, student performance, and administration decisions [23]. Data-driven knowledge discovery is a paradigm that can be applied to DM [24–26]. DM is a multifaceted area that encompasses a variety of topics such as statistics, AI, information technology, learning, data visualization, and retrieval [27].

Because of the enhanced mining application, the educational system has become more balanced [28]. In the sphere of education, the concept of educational data mining (EDM) has rapidly evolved in relation to many types of educational organizations [29]. Further, an academic analyst has been linked to institutional efficiency and student performance issues [30–39]. The EDM [40] encompasses all aspects that have a direct impact on students at the school. As indicated in Table 1, we give a review of the earlier ten researches that used DM techniques in educational settings, extending from 2016 to 2021.

2.1 DM in educational informational management system

EDM is one of the parts of DM, and its main focus is on constructing systems for obtaining hidden information from records of students, which may then be used to improve students' academic performance. Raw data collected from several educational organizations can be transformed into important information, which is used by students, their parents, instructors, educational software developers, and educational researchers in the process of EDM. It can also be viewed of as a system that is a part of the current education system and is capable of generating beneficial interactions with various elements of it. This will allow it to eventually achieve its goal of improving education [59]. EDM is described as the use of standard DM techniques to educational data processing to solve problems in the field of education [60]. There are some examples of EDM applications such as the construction of systems based on e-learning technology [60,61], the educational data clustering [62], and the prediction of student performance [63]. In educational DM, there are several techniques that fall into the categories such as association rule analysis (ARA), sequential pattern, prediction, clustering, classification, and machine learning systems.

2.2 Educational settings based on DM techniques

Clustering [23,41,64–70], classification [44,64,71–73], sequential pattern [67,74,75], prediction [42,67], ARA [44,64], machine learning [76], and ANN [77] are the most well-known DM techniques. From 1995 to 2005, the bulk of studies on educational DM used the ARA technique [78] because it required less knowledge than other techniques [68]. Nonetheless, by the beginning of 2005, the tendency had shifted, and academics

Table 1: Related studies on educational DM

Author	Problem definition	DM technique	Methodology
Durairaj and Vijitha [41]	Student's performance prediction	Classification	DT
Francis and Babu [42]	Student's academic performance prediction	Clustering and classification	NN, support vector machine, DT, and naive bayes (NB)
Akram et al. [43]	Student's academic procrastination prediction	Clustering and classification	Random forest (RF), <i>k</i> -means, NBTree, decision stump, prism, PART, iterative dichotomiser 3 (ID3), ZeroR, open source Java implementation (J48), and One rule (Data mining map) (OneR)
Rojanavas [44]	How can DM aid the admissions process, and how can DM predict a student's future employment?	Classification and association rules	ID3 and DT
Sana et al. [45]	Student's performance prediction	Regression, classification and clustering	Artificial neural network (ANN), DT, and NB
Bharara et al. [46]	Analysis of student's disposition	Clustering	<i>k</i> -means
Nurhayati et al. [47]	Student's graduation prediction	Classification	Multilayer perception (MLP)
Rao et al. [48]	Prediction of students placement	Regression, classification and clustering	NB, RF, J48, NN, random tree (RT), binomial logistic regression, multiple linear regression, regression tree, recursive partitioning, and conditional inference tree
Okubo et al. [49]	Predicting final grades of students	Multiple regression analysis (MRA)	Recurrent neural network
Almarabeh [50]	Student's performance prediction	Classification	NB, ID3, NN, J48, and Bayesian network
Alban and Mauricio [51]	Predicting the university dropouts	MRA	Radial basis function and MLP
Feng [52]	Student's academic performance prediction	Classification	NN and DT
Alshangiti and Namoun [53]	Student's performance prediction	Regression and classification	Fuzzy set rules, lasso linear regression and collaborative filtering (CF)
Hasan et al. [54]	Student's performance prediction in higher educational	Classification	CN@ rule inducer, genetic algorithm, and RF
Tsiakmaki et al. [55]	Student's performance prediction	Classification	Deep neural network-transfer learning
Tadayon and Pottie [56]	Student's performance prediction in educational game	Classification	Hidden Markov model
Lin et al. [57]	Student's performance prediction in educational experiment	Classification	Natural language processing and CART-DT
Saleh et al. [58]	Student's performance prediction in higher educational in Libyan	Clustering and classification	<i>k</i> -means and J84

were increasingly using clustering and classification approaches for analysis [79]. It is usual to generate a set of outputs for an association rule, the bulk of which are uninteresting and difficult to interpret for non-DM users [80]. Researchers must first create the data and check that it is consistent with the desired output before selecting the optimal algorithms [79]. Because data splitting is not required in this procedure, they can use the clustering approach instead of the classification approach when their inquiry is of a modest scale [79]. Furthermore, using the same database as in an earlier study [67], the researchers can always evaluate with alternative algorithms. This would make it easier to see if the same outcomes might be achieved with a different approach.

3 Methods of DM in UIMS

The number of students in schools has expanded dramatically as a result of the continual expansion of enrollment in universities and colleges, considerably increasing the work burden of the management staff in all parts of schools. DM technology is now widely used in a variety of industries, particularly in college and university teaching management systems, where it helps to grasp basic information of students, learning characteristics of master students' and to set up courses of teaching efficiently. Traditional management approach based on manual practice is no longer capable of meeting the demands of today's job. This management approach has a number of flaws, including inefficiency and lack of confidentiality. Furthermore, a great amount of data and files will be produced throughout time, posing significant challenges in terms of searching, updating, and managing them. The UIMS is a critical component of every educational institution, and its content is critical to school managers and decision makers. As a result, the UIMS model should give enough information and query choices for users.

3.1 DM technique

The use of DM technology in the modern management education has a wide variety of applications, and it is now the most important effort and goal in the development of teaching management in Chinese institutions. Instructors should pay special attention to students' physiological and psychological characteristics and then fully engage students in their dominant role before fully engaging teachers in their auxiliary and leading roles. Some information is haphazard if it is done by hand, and it may result in making workers' jobs more difficult and the data more jumbled. Data should be extracted by users so that they can be used more effectively in people's lives. DM technology extracts the data information needed for teaching from a large volume of geographic data based on a geographical database using identification technology and statistical methodologies. By constantly evaluating and identifying these data information, the most practical implementation of data processing method is eventually discovered, providing scientific data assistance for leaders of universities and colleges and functional departments of teaching management.

As DM and knowledge discovery research has progressed, it has created three powerful technical pillars: mathematical statistics, database, and AI. Basic theory, reuse and maintenance of discovery knowledge, data warehouse, discovery algorithm, visualization technology, quantitative and qualitative exchange model, knowledge representation method, knowledge discovery in unstructured and semistructured data, and online DM are currently the main research topics of data mining and knowledge discovery [20]. The most common types of knowledge discovered by DM are as follows:

3.1.1 Generalization knowledge

Generalized descriptive knowledge of category features is referred to as generalized knowledge. According to the microscopic characteristics of data, it discovers the knowledge represented by it, with universality, higher level concept, medium view and macro view, which reflects the common nature of similar things and is the generalization, refinement, and abstraction of data.

3.1.2 Association

Association expresses the understanding of interdependencies or links between events. When two or more qualities are linked, the value of one can be predicted based on the values of the others. There are two phases to discovering the most well-known association rules. The first phase is to iteratively identify all frequent item sets, with the support rate for frequent item sets having to be at least as high as the user's

lowest value. The first phase is to build rules from often occurring item sets with a believability equal to or greater than the user's lowest value. The heart of the association rule discovery algorithm is to identify or find all frequent item sets. The heart of the association rule discovery method is the identification or discovery of all frequent item sets, which also happens to be the section with the most calculations [81].

3.2 Introduction to DM process

The general contents of each step in the DM process are as follows:

3.2.1 Determine business objects

The first stage in DM is to clearly define the business challenge and to comprehend the goal of DM. Although the eventual structure of mining is unknown, the issues that must be addressed should be anticipated. DM is deafening blind and needs to determine on a case by case basis when being deaf or blind have an affect on your ability to make a perception check.

3.2.2 Data preparation

- **Data selection:** By searching all external and internal data information linked to business items, data suitable for DM applications were selected.
- **Data preprocessing:** The data quality was examined to prepare for progressive analysis. Also, the type of mining activity to be carried out was decided.
- **Data transformation:** Data were transformed into a model for analysis. The mining algorithm is built up in this analytical model. It is the key to the success of DM to build an analysis model that is really appropriate for mining algorithm.

3.2.3 DM

Mining the obtained transformed data. All the work can be done automatically except to perfect and select the appropriate mining algorithm.

4 Technological process of UIMS based on DM technology

The technical process of UIMS based on DM technology is described below:

4.1 Application process of association rules in teacher information management module

4.1.1 Data discretization

There are professor system and research series in the title name. Considering the convenience and equivalence of title in DM, title – law is discretized into “senior, associate senior, intermediate, and assistant”

degrees. Undergraduate and lower degrees, master, and doctor are available. The quality of published articles can be classified into science citation index, core, and general [82].

Program is prepared in Java language, and table is created by Access data table. Access is used to create data tables for storing titles, degrees, and published articles. The table structure is shown in Table 1.

4.1.2 Search for frequent item set algorithm

The algorithm of searching frequent predicate set on data cube based on Apriori algorithm is called AprioriCube algorithm. The difference between the algorithm and Apriori algorithm is the calculation of predicate set support. In multidimensional association rule mining based on data cube, a predicate set is a combination of dimension members of different dimensions of data cube ($d_1...d_n|count$), and the support count of predicate set is the frequent measure value stored in cube squares.

4.1.3 Find frequent item sets

A one-dimensional table is created where y_1 represents the attribute of title or degree and sup represents support. A table is created where y_1 and y_2 , represent the attributes of title or degree and sup represents support. A three-dimensional table is created where y_1 and y_2 , represent the attributes of title or degree and sup represents support. All one-dimensional member inputs are deleted from infrequent item sets, minsup is the minimum supported frequency, and RECC is the number of records [83].

4.2 Apriori algorithm is used to determine the realization process of the correlation between the courses selected by students

4.2.1 Objectives

According to the data of students' course selection and preselection in the database of students' course selection management, the association rules between courses are mined to determine the association relation of students' course selection, which provides the basis for propaganda planning and course classification.

4.2.2 Determine the type of DM

Assuming that global is a collection of courses available from distance education providers, each course has a Boolean variable demonstrating the absence or presence of the course. Each course selection sequence can be represented by a Boolean vector. Boolean vector can be analyzed to obtain the pattern of course selection reflecting the frequent association of courses. These patterns can be expressed in the form of association rules. Therefore, it can be determined that to find the association relationship between the courses selected by students, we can mine the association rules of selected courses in the student elective management database. Because only the one-dimensional data of the courses selected by students need to be considered, the method adopted in this article is as follows: First, the Apriori algorithm 10 is used to find frequent item sets, and then frequent item sets are generated to generate association rules [84].

4.2.3 Process

The process of UIMS is explained in Algorithm 1.

Algorithm 1

- Step 1:** Determine the target data of DM – the data of students’ elective courses and preselected courses – in the student elective management database, including course names and learning seasons.
- Step 2:** Collect task-related data sets based on the following relational query.
- Step 3:** Determine the minimum support threshold minsup.
- Step 4:** Use Apriori algorithm to discover the frequent itemsets.
- Step 5:** Make association rules from the frequent itemsets.
-

Assume that the number of tuples selected is 9, that is, $|D| = 9$. The tuple identifier is represented by TID and stored in lexicographical order, as shown in Figure 1.

Figure 1: Set of related tuples.

5 Result and experimental test

Courses in all areas: {I1 = computer graphics, I2 = “image processing,” I3 = “computer-aided design,” I4 = “IC Computer-Aided Design (ICCAD) software tools,” I5 = “semiconductor theory,” I6 = “large-scale analog integrated circuits,” I7 = “accounting,” I8 = “computer fundamentals,” I9 = “architectural design”}. D is obtained through relational query from the student course selection management database, as shown in Figure 2. Let the minimum transaction support count be 2, $\min_{\text{conf}} = 70\%$, then the output rule is: $I3 > I4$, $I9 > I13$; that is, students who choose the course of “computer-aided design” also tend to choose the course of “ICCAD software tools,” and students who choose the course of “architectural design” also tend to choose the course of “computer-aided design.”

Figure 2: Set of related tuples of student course selection data.

5.1 The design and improvement of Apriori algorithm based on array is applied in the analysis of test scores

Because of Boolean variable, first need is to convert Boolean variables to unidimensional values with early grades as an example, this study divides the preliminary achievement good, medium, and poor 3, respectively, for I1, I2 I3, sex to male with I2, said woman with I5, said the work at ordinary times is also divided into good, poor, in three levels, I6, I7, and I8 are used to represent the class performance, which is also divided into three grades: good, medium and poor; I9, I10, and I11 are used to represent the rewards and punishments, which are classified as I12, I13, and I14. The final scores are divided into excellent (above 85 points), good (70–84), and passing (60–69). Moreover, fail (below 60 points) are represented by I15, I16, I17, and I18, respectively. In this case, 1 represents the yes of a Boolean variable and 0 represents the no of a Boolean variable, which has the advantage of facilitating array operations. It can be seen from the above that given a student whose early score is medium, gender is male, daily homework is good, class performance is good, rewards and punishments are not good, and the final score is good, it can be represented as 010101001000100100. Here, the early score is the average of previous comprehensive score, 80 points or more is good. The program can realize the preprocessing of data and read the processed data into the array. Table 2 is a view of the data converted to Booleans [85] (Table 3).

Table 2: Data table of titles, degrees, and published articles

Fields that	The field name	The field type	Field constraint	The default value
The title	zc	Character	A primary key	There is no
A degree in	xw	Character		There is no
The paper	1w	Character		There is no

Here are the specific steps for DM:

- 1) The occurrence times of each subitem are counted, the support degree of each subitem is calculated, the items meeting the minimum support degree into the dimensional array P1 are stored, that is, frequent 1

Table 3: Data converted to Booleans

Xh	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18
1	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0
2	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
3	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1	0	0
4	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	1	0	0
5	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	0	1
6	1	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	0
7	0	0	1	0	1	1	0	0	1	0	0	0	0	1	0	1	0	0
8	0	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0	1	0
9	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1
10	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0	1	0	0

item set is generated, and the items that do not meet the minimum support degree are deleted. That is, the size of a two-dimensional array is compressed.

- 2) The subsequent frequent item sets are divided into two steps. Candidate items are generated in the first step, and frequent item sets are generated in the second step. The specific process is as follows: First, by frequent $(n - 1)$ item set $(N \geq 2)$, each item is naturally connected to generate the candidate item set of N , which is stored in the array P_n in an ascending order. Second, the support degree of candidate N is obtained by scanning the two-dimensional array and recorded in the array P_n . Finally, delete rows and columns whose support count is less than the minimum support $\min_{\text{sup}}$. Until all frequent item sets are found, if the number of a candidate item set is zero, the operation is stopped. Finally, the frequent set for all items is printed. The number of times the program scans the transaction array depends on the maximum length of the frequent item set.
- 3) The confidence of each nonempty subset of the final frequent item set is calculated, the records less than the minimum confidence threshold are deleted, and finally rules are generated.

University management or teachers can use management information systems to identify special pupils and give the foundation for creating necessary psychological education to be successful in the field of education. In the evolution of informatization of management in universities and colleges, DM is an unavoidable stage. Universities must use electronic information technology to increase the management efficiency and quality, establish a perfect education IMS, and perform in-depth analyses of resources of campus and data utilizing modern DM approaches to enhance informationization of management. To properly promote the management informatization development in universities, the educational or research management system is combined with the authentic state of schools, fully understand the bottleneck issues in DM, improve the precision and high quality of data analysis, and fully realize the true role of DM [86].

In relation to information, it is critical to build and improve a UIMS with a greater functional level, to utilize advanced and scientific tools of statistical analysis, and to utilize progressive and scientific mining expertise to deeply analyze management data, in order to improve management intensity of intelligence [87–89].

To boost management intelligence intensity, it is necessary to create and improve a UIMS with a greater functional level, to apply advanced and scientific statistical analytical techniques, and to use scientific and advanced mining technology to thoroughly examine management data.

6 Conclusion

This study presents the application research of DM in the development of university IMS. First, it introduces DM and its process, and studies the application of association rules in UIMS. It shows that the curriculum

covers the whole field: {I1 = computer graphics, I2 = “image processing,” I3 = “computer-aided design,” I4 = “ICCAD software tools,” I5 = “semiconductor theory,” I6 = “large-scale analog integrated circuits,” I7 = “accounting,” I8 = “computer fundamentals,” I9 = “architectural design”}. D is obtained through relational query from the student course selection management database, as shown in Figure 2. Let the minimum transaction support count be 2, then the output rule is, that is, students who choose the course of “computer-aided design” also tend to choose the course of “ICCAD software tools,” and students who choose the course of “architectural design” also tend to choose the course of “computer-aided design.” The rational procedure of the conclusion of this method has a great dependence on data. When the amount of historical data is increasing and the curriculum planning is improving, it will form a virtuous circle to promote the rationality of the combination of curriculum modules. The application of DM theory in university information database can be further deepened, and the relevant algorithms of DM can be optimized according to the characteristics of university information database. To play a greater role in talent evaluation, discipline echelon construction, postmanagement and the formulation of management information policy, DM is carried out from different angles such as multilayer and multidimension. In future, different machine learning techniques will be used to develop UIMS.

Conflict of interest: Authors state no conflict of interest.

References

- [1] Zhang Y, Lu W-X, Guo J-Y, Zhao H-Q, Yang Q-C, Chen M. Geo-environmental impact assessment and management information system for the mining area, northeast China. *Environ Earth Sci.* 2015;74(10):7173–85.
- [2] Haizhao G. Research on university educational administration management based on data mining technology. *Digital World.* 2020;176(6):213–3.
- [3] Nazarenko MA, Tatiana VK. Big data in modern higher education. Benefits and criticism. 2017 International Conference “Quality Management, Transport and Information Security, Information Technologies” (IT&QM&IS). IEEE; 2017. p. 676–9.
- [4] Qian Z, Rui F. Overview of the application of data mining in the information management of universities and colleges. *Data Min.* 2019;9(1):1–7.
- [5] Kaiwei Z. Research on the application of data mining technology in university educational administration. *China Manag Inf.* 2018;21(4):123–4.
- [6] Jach T, Magiera E, Froelich W. Application of HADOOP to store and process big data gathered from an urban water distribution system. *Proc Eng.* 2015;119:1375–80.
- [7] Liu X, Ma Y, Huang G, Zhao J, Mei H, Liu Y. Data-driven composition for service-oriented situational web applications. *IEEE Trans Serv Comput.* 2014;8(1):2–16.
- [8] Liang J. Research on education management and decision optimization based on cloud computing from the perspective of big data. In 2021 4th International Conference on Information Systems and Computer Aided Education; 2021. p. 404–7.
- [9] Cerone A. Model mining. *J Intell Inf Syst.* 2019;52(3):501–32.
- [10] Garg L, McClean S, Meenan B, Millard P. Non-homogeneous Markov models for sequential pattern mining of healthcare data. *IMA J Manag Mathematics.* 2009;20(4):327–44.
- [11] Ping. Z. Challenges and countermeasures of higher education management in the era of big data. *J Wuyi Univ.* 2017;36(10):92–5.
- [12] Ronghui G, Gaoda H. The value orientation and realization path of higher education reform in the era of big data. *China Audio-v Educ.* 2015;346(11):70–6.
- [13] Zheng L. Application of big data mining technology in online education platform. *China Manag Informationization.* 2017;20(16):223–4.
- [14] Quan S, Yuan Q, Ling S. Research on online learning process supervision based on educational data mining. *Mod Educ Technol.* 2016;26(6):87–93.
- [15] Zhengqing W, Feng DJ. The process and guarantee of education data quality management in the United States in the era of big data. *Mod Distance Educ Res.* 2019;31(5):96–103, 112.
- [16] Mago VK, Giabbanelli PJ. Special issue on intelligent healthcare systems. *J Intell Syst.* 2016;25(1):1–2.
- [17] Li A, Liu K, Ge Z. Application of data mining in the colleges’ in-class teaching quality evaluation system. *J Comput.* 2015;10(3):166–75.
- [18] Yanhao H. Study on the application of electric power big data technology in power system simulation. *Proc Csee.* 2015;1:256–6.

- [19] Bostani H, Sheikhan M. Hybrid of binary gravitational search algorithm and mutual information for feature selection in intrusion detection systems. *Soft Comput.* 2017;21(9):2307–24.
- [20] Lee HP, Hong D-P, Han E, Kim SH, Yun I. Analysis of the characteristics of expressway traffic information propagation using Twitter. *KSCE J Civ Eng.* 2016;20(7):2587–97.
- [21] Salloum SA, Al-Emran M, Shaalan K. Mining social media text: extracting knowledge from Facebook. *Int J Comput Digital Syst.* 2017;6(2):73–81.
- [22] Salloum SA, Mhamdi C, Al-Emran M, Shaalan K. Analysis and classification of Arabic newspapers' Facebook pages using text mining techniques. *Int J Inf Technol Lang Stud.* 2017;1(2):8–17.
- [23] Fernandes E, Holanda M, Victorino M, Borges V, Carvalho R, Van Erven G. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *J Bus Res.* 2019;94:335–43.
- [24] Cummins MR. Nonhypothesis-driven research: data mining and knowledge discovery. *Clinical research informatics.* Cham: Springer; 2019. p. 341–56.
- [25] Salloum SA, Al-Emran M, Abdel Monem A, Shaalan K. A survey of text mining in social media: Facebook and Twitter perspectives. *Adv Sci Technol Eng Syst J.* 2017;2(1):127–33.
- [26] Alomari KM, AlHamad AQ, Salloum S, Salloum SA. Prediction of the digital game rating systems based on the ESRB. *Opcion.* 2019;35(19):1368–93.
- [27] Arunachalam AS, Velmurugan T. Analyzing student performance using evolutionary artificial neural network algorithm. *Int J Eng & Technol.* 2018;7(2.26):67–73.
- [28] Romero C, Ventura S, García E. Data mining in course management systems: moodle case study and tutorial. *Computers Educ.* 2008;51(1):368–84.
- [29] Sachin RB, Shelake Vijay M. A survey and future vision of data mining in educational field. 2012 Second International Conference on Advanced Computing & Communication Technologies. IEEE; 2012. p. 96–100.
- [30] Salloum SA, Alhamad AQM, Al-Emran M, Monem AA, Shaalan K. Exploring students' acceptance of e-learning through the development of a comprehensive technology acceptance model. *IEEE Access.* 2019;7:128445–62.
- [31] Alshurideh M, Salloum SA, Kurdi BA, Al-Emran M. Factors affecting the social networks acceptance: an empirical study using PLS-SEM approach. *Proceedings of the 2019 8th International Conference on Software and Computer Applications;* 2019. p. 414–8.
- [32] Salloum SA, Shaalan K. Factors affecting students' acceptance of e-learning system in higher education using UTAUT and structural equation modeling approaches. *International Conference on Advanced Intelligent Systems and Informatics;* 2018. p. 469–80.
- [33] Salloum SA, Al-Emran M, Habes M, Alghizzawi M, Ghani MA, Shaalan K. Understanding the impact of social media practices on e-learning systems acceptance. *International Conference on Advanced Intelligent Systems and Informatics;* 2019. p. 360–9.
- [34] Salloum SA, Mhamdi C, Al Kurdi B, Shaalan K. Factors affecting the adoption and meaningful use of social media: a structural equation modeling approach. *Int J Inf Technol Lang Stud.* 2018;2(3):96–109.
- [35] Salloum SA, Maqableh W, Mhamdi C, Al Kurdi B, Shaalan K. Studying the social media adoption by university students in the United Arab Emirates. *Int J Inf Technol Lang Stud.* 2018;2(3):83–95.
- [36] Salloum SAS, Shaalan K. Investigating students' acceptance of e-learning system in higher educational environments in the UAE: applying the extended technology acceptance model (TAM). *The British University in Dubai;* 2018.
- [37] Habes M, Alghizzawi M, Khalaf R, Salloum SA, Ghani MA. The relationship between social media and academic performance: Facebook perspective. *Int J Inf Technol Lang Stud.* 2018;2(1):12–8.
- [38] Salloum SA, Al-Emran M, Shaalan K, Tarhini A. Factors affecting the E-learning acceptance: a case study from UAE. *Educ Inf Technol.* 2019;24:509–30.
- [39] Al-Emran M, Salloum SA. Students' attitudes towards the use of mobile technologies in e-evaluation. *Int J Interact Mob Technol.* 2017;11(5):195–202.
- [40] Kabakchieva D. Predicting student performance by using data mining methods for classification. *Cybern Inf Technol.* 2013;13(1):61–72.
- [41] Durairaj M, Vijitha C. Educational data mining for prediction of student performance using clustering algorithms. *Int J Comput Sci Inf Technol.* 2014;5(4):5987–91.
- [42] Francis BK, Babu SS. Predicting academic performance of students using a hybrid data mining approach. *J Med Syst.* 2019;43(6):162.
- [43] Akram A, Fu C, Li Y, Javed MY, Lin R, Jiang Y, et al. Predicting students' academic procrastination in blended learning course using homework submission data. *IEEE Access.* 2019;7:102487–98.
- [44] Rojanavas P. Educational data analytics using association rule mining and classification. 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON); 2019. p. 142–5.
- [45] Sana B, Siddiqui IF, Arain QA. Analyzing students' academic performance through educational data mining. *3c Tecnol glosas innovación Apl a la pyme.* 2019;8(29):402–21.
- [46] Bharara S, Sabitha S, Bansal A. Application of learning analytics using clustering data Mining for Students' disposition analysis. *Educ Inf Technol.* 2018;23(2):957–84.

- [47] Nurhayati OD, Bachri OS, Supriyanto A, Hasbullah M. Graduation prediction system using artificial neural network. *Int J Mech Eng Technol*. 2018;9(7):1051–7.
- [48] Rao KS, Swapna N, Kumar PP. Educational data mining for student placement prediction using machine learning algorithms. *Int J Eng Technol Sci*. 2018;7(1.2):43–6.
- [49] Okubo F, Yamashita T, Shimada A, Ogata H. A neural network approach for students' performance prediction. *LAK*. 2017;2017:598–9.
- [50] Almarabeh H. Analysis of students' performance by using different data mining classifiers. *Int J Mod Educ Comput Sci*. 2017;9(8):9–15.
- [51] Alban M, Mauricio D. Neural networks to predict dropout at the universities. *Int J Mach Learn Comput*. 2019;9(2):149–53.
- [52] Feng J. Predicting students' academic performance with decision tree and neural network. PhD Dissertation. 2019.
- [53] Alshanqiti A, Namoun A. Predicting student performance and its influential factors using hybrid regression and multi-label classification. *IEEE Access*. 2020;8:203827–44.
- [54] Hasan R, Palaniappan S, Mahmood S, Abbas A, Sarker KU, Sattar MU. Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Appl Sci*. 2020;10(11):3894.
- [55] Tsiakmaki M, Kostopoulos G, Kotsiantis S, Ragos O. Transfer learning from deep neural networks for predicting student performance. *Appl Sci*. 2020;10(6):2145.
- [56] Tadayon M, Pottie GJ. Predicting student performance in an educational game using a hidden markov model. *IEEE Trans Educ*. 2020;63(4):299–304.
- [57] Lin L, Tan LWL, Kan NHL, Tan OK, Sze CC, Goh WWB. Predicting student performance in experiential education. *International Conference on Database and Expert Systems Applications*. Cham: Springer; 2021. p. 328–334.
- [58] Saleh MA, Palaniappan S, Abdallah NAA, Baroud MMB. Predicting student performance using data mining and learning analysis technique in Libyan Higher Education. *Periodicals Eng Nat Sci (PEN)*. 2021;9(3):509–19.
- [59] Jie W, Hai-yan L, Biao C, Yuan Z. Application of educational data mining on analysis of students' online learning behavior. *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*; 2017. p. 1011–5.
- [60] Baker RSJD, Yacef K. The state of educational data mining in 2009: a review and future visions. *JEDM: J Educ Data Min*. 2019;1(1):3–17.
- [61] Lara JA, Lizcano D, Martínez MA, Pazos J, Riera T. A system for knowledge discovery in e-learning environments within the European higher education area-application to student data from open university of madrid, UDIMA. *Comput Educ*. 2014;72:23–36.
- [62] Chakraborty B, Chakma K, Mukherjee A. A density-based clustering algorithm and experiments on student dataset with noises using Rough set theory. *2016 IEEE International Conference on Engineering and Technology (ICETECH)*; 2016. p. 431–6.
- [63] Chauhan N, Shah K, Karn D, Dalal J. Prediction of student's performance using machine learning. In *2nd International Conference on Advances in Science & Technology (ICAST)*; 2019.
- [64] Pechenizkiy M, Calders T, Vasilyeva E, De Bra P. Mining the student assessment data: lessons drawn from a small scale case study. *The 1st International Conference on Educational Data Mining*; 2008.
- [65] Shih Y-C, Huang P-R, Hsu Y-C, Chen SY. A complete understanding of disorientation problems in Web-based learning. *Turkish Online J Educ Technol*. 2012;11(3):1–13.
- [66] Talavera, L, Gaudioso, E. Mining student data to characterize similar behavior groups in unstructured collaboration spaces. *Workshop on Artificial Intelligence in CSCL*. 16th European Conference on Artificial Intelligence; 2004. p. 17–23.
- [67] Perera D, Kay J, Koprinska I, Yacef K, Zaiane OR. Clustering and sequential pattern mining of online collaborative learning data. *IEEE Trans Knowl Data Eng*. 2008;21(6):759–72.
- [68] Dutt A, Aghabozrgi S, Ismail MAB, Mahrooian H. Clustering algorithms applied in educational data mining. *Int J Inf Electron Eng*. 2015;5(2):112.
- [69] Bogarín A, Romero C, Cerezo R, Sánchez-Santillán M. Clustering for improving educational process mining. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*; 2014. p. 11–5.
- [70] Palomo-Duarte M, Berns A, Yañez Escolano A, Doderio J-M. Clustering analysis of game-based learning: worth it for all students? *J Gaming Virtual Worlds*. 2019;11(1):45–66.
- [71] Ahmed ABED, Elaraby IS. Data mining: a prediction for student's performance using classification method. *World J Comput Appl Technol*. 2014;2(2):43–7.
- [72] Anjewierden A, Kolloffel B, Hulshof C. Towards educational data mining: using data mining methods for automated chat analysis to understand and support inquiry learning processes. In *International Workshop on Applying Data Mining in e-Learning (ADML 2007)*; 2007. p. 27–36.
- [73] Adebayo AO, Chaubey MS. Data mining classification techniques on the analysis of student's performance. *GSJ*. 2019;7(4):45–52.
- [74] Keskin S, Şahin M, Yurdugül H. Online learners' navigational patterns based on data mining in terms of learning achievement. *Learning technologies for transforming large-scale teaching, learning, and assessment*. Springer; 2019. p. 105–21.

- [75] Kay J, Maisonneuve N, Yacef K, Zaiane O. Mining patterns of events in students' teamwork data. *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (ITS 2006)*; 2006. p. 45–52.
- [76] Kučak D, Juričić V, Đambić G. Machine learning in education-a survey of current research trends. *Ann DAAAM and Proc.* 2018;29:406–10.
- [77] Coelho OB, Silveira I. Deep learning applied to learning analytics and educational data mining: a systematic literature review. *Brazilian Symp Comput Educ (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2017;28(1):143.
- [78] Romero C, Ventura S. Educational data mining: a survey from 1995 to 2005. *Expert Syst Appl.* 2007;33(1):135–46.
- [79] Salloom SA, Alshurideh M, Elnagar A, Shaalan K. Mining in educational data: review and future directions. *AICV*. 2020;92–102.
- [80] García E, Romero C, Ventura S, Calders T. Drawbacks and solutions of applying association rule mining in learning management systems. *Proceedings of the International Workshop on Applying Data Mining in e-Learning (ADML 2007)*, Crete, Greece; 2007. p. 13–22.
- [81] Fu K, Murphy J, Yang M, Otto K, Jensen D, Wood K. Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement. *Res Eng Des.* 2015;26(1):77–95.
- [82] Blar N, Jafar FA. A survey on the application of robotic teacher in Malaysia. *Advanced Computer and Communication Engineering Technology*. Cham: Springer; 2015. p. 605–13.
- [83] Holimchayachotikul P, Leksakul K. Predictive performance measurement system for retail industry using neuro-fuzzy system based on swarm intelligence. *Soft Comput.* 2017;21(7):1895–912.
- [84] Cai Y, Wang J, Chen Y, Wang T, Tian H, Luo W. Adaptive direction information in differential evolution for numerical optimization. *Soft Comput.* 2016;20(2):465–94.
- [85] Choy KL, Ho GTS, Lee CKH. A RFID-based storage assignment system for enhancing the efficiency of order picking. *J Intell Manuf.* 2017;28(1):111–29.
- [86] Balyan V. Outage probability of cognitive radio network utilizing non orthogonal multiple access. *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE; 2020. p. 751–5.
- [87] Kaushik M, Gupta SH, Balyan V. Evaluating threshold distance by using eigen values and analyzing its impact on the performance of WBAN. *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE; 2019. p. 864–7.
- [88] Balyan V, Daniels R. Resource allocation for NOMA based networks using relays: cell centre and cell edge users. *Int J Smart Sens Intell Syst.* 2020;13(1):1–18.
- [89] Sharma A, Singh PK, Sharma A, Kumar R. An efficient architecture for the accurate detection and monitoring of an event through the sky. *Comput Commun.* 2019;148:115–28.