Education is An Overview of Data Mining and The Ability to Predict the Performance of Students

Mahjouba Ali Saleh✉, Sellappan Palaniappan, Nasaraldeen Ali Alghazali Abdalla

[1]University of Science and Technology Kuala Lumpur, Malaysia
[2]Bani Waleed University Libya

Info Articles

Abstract

This research provides a review of the state of the art with respect to EDM and discusses the most relevant work in this area to date. Each study has been discussed considering type of data and data mining techniques used, and the kind of the educational task that they resolve. EDM is upcoming research area related to well-established areas of research such as e- learning, tutoring systems, web mining, data mining. Current literature show how fast educational data analysis area is growing and there is an increasing number of contributions that publish in International Journals and Conferences every year. However, educational data mining is still not a mature area. Some interesting future suggestion to develop this area has been presented. This research is a presentation of current and ancient literature of Predicting Student Performance using Data Mining.

*Alamat Korespondensi:
E-mail: mahjouba.ali@phd.must.edu.my

INTRODUCTION

In the context of educational institutions, the quality of teaching process is the ability to meet students and markets' needs. The concept of quality of education institutions indicates provide the services that satisfy of students, academic staff, and other participants in the education system. Therefore, these institutions seek to contribute to improve the quality of education through create of an educated human capital. In this context, education institutions collect data and information on a regular basis, which create an enormous amount of data that can be analyzed and integrated and then converting into valuable knowledge to support students, educators, administration and community.

Accordingly, a new area of research has emerged that called educational data mining (EDM) (Qasem & Nagi, 2014). In the recent years Educational Data Mining (EDM) got popularity as an emerging field within Data mining. This is because of the growth of educational software and Internet usage for educational purposes. According to(Costa, Fonseca, Santana, de Araújo, & Rego, 2017), EDM produces a big amount of data that can be used to predict student's learning level. To understand educational issues, EDM is used for analyzing educational data through takes advantages of statistical, information technology, machine-learning, data- mining algorithms, artificial intelligence, and database management system (Ahmed, Rizaner, & Ulusoy, 2016). Although, it is still relatively new field, educational data mining techniques has been successfully implemented in the educational context. The objective of EDM is to identify and extract of new and potentially valuable hidden knowledge from the data. EDM techniques intended to develop a model that can derive the conclusions on students' academic success. However, predicting students' academic performance becomes more challenging due to the large volume of data in educational databases (Asif, Merceron, Ali, & Haider, 2017) divided EDM objective into two objectives that are: (a)   academic objectives that can be oriented by different fields such as person oriented, department/institutions oriented or domain oriented. (b) administrative objectives that can be administrator oriented. In Libya.

Many freshmen students are not prepared to make a successful shift from primary school to high school and also may be underprepared to face several challenges in high school, in which courses move to an advance level of knowledge, which can be very stressful. However, there is lack of research attention to use EDM techniques to address the student performance and progress in the Libyan context. In Libya, predicting students' performance becomes more challenging due to (a) the large volume of data in educational databases; (b) lack of any electronic system to store the students' records; (c) lack of system to analyze and monitor the student performance and progress.

Existing system is traditional and does not involve any analysis process of students' learning. Libyan educational system does not deal with dropouts' cases, does not caution the students about the deficiency in attendance, does not identify the weak student and inform the teachers. Therefore, existing educational system in Libyan institutions is still primitive and unable to identify the most suitable methods for solving the educational issues. Specially, unable to predict and improve the students' performance due to the lack of investigations on the factors affecting students' performance.

Although, there are abundance studies in the literature that have discussed the mining of educational data, and there is agreement about the importance of data mining in the educational context; using data mining to enhance the education system is still relatively new. Existing research is too specific and most of the current research is limited to a context.

There is very limited research that attempts to identify factors that affect students' performance. There are few researches that attempted to create a standard prediction process for student performance and that can be widely used. This concern has motivated the undertaking of the

present research. In the Libyan context, predicting students' performance becomes more challenging due to several reasons as following: (a) the large volume of data in educational databases; (b) lack of system to analyze and monitor the student performance and progress.

Existing system does not involve any prediction about pass or fail percentage based on students' performance, does not deal with dropouts, does not caution the students about the deficiency in attendance, does not identify the weak student and inform the teachers. Existing prediction methods in Libyan institutions is still insufficient to identify the most suitable methods for predicting the students' performance due to the lack of investigations on the factors affecting students' performance (Yang & Li, 2018). To fill this research gap, there is a need to further research to identify and evaluate factors that may influence the students' performance (Rodrigues, Zárate & Isotani, 2018).

It is important to conduct a future research on EDM to deepen our understanding of possibility to predict students' performance, doing such prediction will help to improve educational outcomes (Asif et al., 2017). Without addressing this gap, it is questionable whether the literature provides us consistent methods for predicting the students' performance, provide a system to analyze and monitor the students' performance and progress to improve the education system quality. Furthermore, in real world, there are some data related to the students' performance in their attendance, quizzes exam, and monthly exams. Therefore, this study also applies the data mining techniques to the students' records data to obtain these types of data.

Current literature provides us many reliable DM techniques that can help to predict students' performance, and to study the correlation between students' attributes and their performance. This research uses EDM techniques i.e., the K-means algorithm for clustering, the J48 algorithm for classification, and the Apriori algorithm for association rules mining experiments to achieve this study objectives. It is argued that EDM techniques i.e., clustering, classification, and association rule mining, can provide more useful results and knowledge about student learning that beyond the statistical analyses (Peral Cortés, Maté, & Marco Such, 2017).

The purpose is to investigate the impacts of these academic practices on Libyan students' performance. This additional focus of this study may or may not directly contribute to the understanding of the impact of learning behavior of performance, but it is important to learn about Libyan students' performance in general. For example, by knowing early at the beginning of the school(s) semester that some students have made bad marks and their progress is unsatisfactory in the course, the educator could offer them an extra lesson or take other corrective actions to help them.

METHODS

Educational data mining is a computational data process that used to extract hidden and potentially useful information and patterns from different perspectives (Dubey, 2016). EDM research aims to provide a deeper understanding of the key factors that impact on students' learning through extracting relevant and important information that has not previously been known that is one of useful and effective features for data mining techniques. It is argued that data mining techniques can be directed to the various acts of the educational process (Badr, Algobail, Almutairi, & Almutery, 2016; Khasanah, 2017). This can be beneficial for students, educators, administration, and community. Thus, support the specific needs of each of the participants in the educational process.

Practically, using the EDM in Libyan institutions will contribute to IT field. This is because it will provide insights and valuable information to instructors, schools managers and educational institutions about the beneficiary of using It tools and EDM techniques to extract information from

the data that extract from the Libyan high schools; and explain how the useful information and patterns can be used in predicting students' performance. Thus, it would assist the educators and school managers in improving teaching approach effectively. Further, educators and school managers could also monitor their students' progress and achievements. Moreover, the prediction process can result important information and knowledge that useful for educators to: (a) receive feedback about students' performance to better understand their students' learning behaviors; (b) help them to continuously evaluate the students' progress and performance. Thus, they can warn the students about the deficiency in attendance and identify the weak student and motivate them to improve their performance.

On the other side, prediction process can help to inform students about how their learning behavior is associated with negative or positive performance/outcomes. It is commonly to use DM techniques to predict future performance based on students' previous academic performance. This task is the aim of data mining techniques, such predictions help students know how well they will do in the future before they register, so, they can avoid poor performance. Moreover, as EDM techniques centers on discovering, detecting, and explaining educational phenomenon. It can greatly help administration to improve the understanding of the teaching process, thus, allowing the administration to improve the systems performance. Furthermore, it will help the instructors and schools' managers to identify students that more likely to fail and then take appropriate actions and change strategies to improve student performance.

DISCUSSION

Studies on educational data mining have increased in recent times (Juhaňák, Zounek, & Rohlíková, 2019). This increase may have resulted from the need for improve the current educational system (Ahmed et al., 2016). The previous studies on using data mining in the educational context have been conducted in three directions: (a) prove the possibility using data mining methods in the educational context; (b) predicting the outcomes of the performance using data mining approaches; (c) evaluation the accuracy of the data mining techniques in the educational context. The following sections review some of these studies.

Using Data Mining Methods in The Educational Context

Current literature shows that data-mining methods has been applied by scholars in the context of education for various purposes and to accomplish various evaluation tasks. Some literature suggests areas for the application of data mining in the educational context, for example, a suggested four major areas for the application of EDM: (a) improving student models; (b) improving the domain model; (c) studying pedagogical support using learning software; (d) scientific research on student learning. Similarly, a suggested five major   tasks/applications for EDM: (a) to evaluate student academic performance; (b) to provide complementary courses according to the student's learning behavior; (c) to evaluate educational resources available on Web courses; (d) to give feedback to teachers and students in distance education courses; and (e) to address atypical behaviors of student learning. divided the major methods of mining educational multiple data sources into four groups, (1) pattern analysis; (2) multiple data source classification; (3) multiple data source clustering; (4) multiple data source fusion.

They also suggest that these approaches need for further research. Rodrigues, Zárate & Isotani (2018) based on literature review found that EDM methods expanded into several areas: (a) monitoring and evaluation of teaching-learning process; (b) administrators' evaluation; (c) learning risks recommendation and recovery of educational media.

In the same vein, some literature proves the possibility using data mining methods in the educational context. For instance, Ismail, and Herawan (2017) based on a systematic literature review on clustering algorithm and its applicability. and usability in the context of educational data mining. They provided future insights and suggested possibilities for further research. Kovanovic, Baker, and Gasevic (2017) discussed the importance of using a data analysis toolbox for the practice of EDM/LA research. Further, Fonseca and Namen (2016) used the KDD methodology with a focus on the data mining stage. Through discuss and analyses of discovered patterns, they identified factors that relate profiles and their influences- positively and negatively on students' Mathematics learning.

Predicting the Student Performance Using Data Mining

Using data mining methods to predict student performance, current literature provides compelling arguments in this context. Ali, and Haider (2017) used EDM to study the performance of undergraduate students. They focused on two aspects of student performance. Predicting students' achievement at the end of a four-year study program, Studying students' progressions and combining them with prediction results (Angeli, Howard, Ma, Yang, & Kirschner, 2017).

They identified two important groups of students: low achieving and high achieving. Their results indicate that using a small number of courses as indicators of good or poor performance will help to provide timely warnings and give support to low performance students and provide advice and opportunities to high performing students. Fernandes et al., (2018) conducted a predictive analysis of the academic performance of students in public schools in Brazil. They performed a descriptive statistical analysis on data that resulted in two datasets.

The first dataset includes variables obtained prior to the start of the school year, and the second contained academic variables collected two months after the semester began. In this study they found that, although the attributes 'grades' and 'absences' were the most relevant for predicting student academic performance, the analysis of demographic attributes reveals that 'neighborhood', 'school' and 'age' are also can be considered as indicators of a student's academic success or failure. Developed a tutoring action plan to predict whether a student will drop out of a course (Burgos et al., 2018).

The suggested tutoring action plan reduced the dropout rate by 14% in e-learning courses with respect to previous academic years in which no dropout prevention mechanism was applied. The CBA rule-generation algorithm to conduct two experiments. In the first experiment they used students' grades in two English courses and two mathematics courses, which generated four rules with accuracy of 62.75%. In the second experiment they used students' grades only in two English courses, generating four rules with accuracy of 67.33% (da Fonseca Silveira, Holanda, de Carvalho Victorino, & Ladeira, 2019).

The study found that students' performance in English courses has a significant predictive effect on their performance in the programming course. In the other side, Kabakchieva, (2013) implemented data mining project at UNWE is following the CRISP- DM (Cross-Industry Standard Process for Data Mining) to examine the high potential of data mining applications for university management. The results show that the selected data mining algorithms that used for classification achieved not remarkable prediction rates vary between 52-67 % (Wang et al., 2018).

Evaluating the Performance of Various Classification Techniques of Data Mining

There are several feature selections to select high influence attributes with student performance, using 104 data of students in academic year 2007 with 12 attributes in Indonesia. The study found that Bayesian Network has higher accuracy rate and outperforming Decision Tree as classification algorithm to know the best prediction result of students' performance (Menon & Islam,

2017). There are two techniques of data mining: association rules mining and fuzzy representations to examine student learning, behaviors, and experiences within computer-supported classroom activities. Two studies have been conducted one in Europe and another in Australia.

The results show that association rules mining is a useful method to collect reliable data about learners' use of the simulation and their performance with it, while Fuzzy representations mining can use for guiding and monitoring school-based technology integration efforts. Costa, et al., (2017) conducted Comparative study on the effectiveness of EDM techniques to early predict students fail in introductory programming courses in Brazil. They found that EDM techniques can early identify students' fail, and the effectiveness of some of these techniques is improved after applying the data pre-processing and/or algorithms fine-tuning (Deeva & De Weerdt, 2018).

The support vector machine technique outperforms the other techniques in a statistically significant way. Course selection is a model that helps students in selection of course, the model uses two of the decision tree classification algorithms: ID3 and J48. They found that J48 achieved a better accuracy performance of 83.75 % when compared to an accuracy of 69.27 % for the ID3 algorithm. Natek and Zwilling (2014) used data mining for small student data sets by comparing two different data mining tools: the MS Excel tool and the WEKA Data Mining tool. They found that both tools show a relative high prediction success and thus strength the conclusion that they can be relevant tool for developing knowledge management systems at the higher education institutions. Further, decision trees algorithms were very practical to work on small data sets (Rizvi, Rienties, Rogaten, & Kizilcec, 2020).

Review of Related EDM Research

Researchers have been found that DM techniques capable of explaining the causes of educational issues such as rates of course failure and dropout and poor academic performance. To improve these rates, many data mining techniques have been applied to the educational data context like clustering, classification, and association rules mining which provided promising results (Nebot, Mugica, & Castro, 2020). The two most predominant categories of EDM methods which encompasses of: (a) Statistics and visualization; (b) Web mining, which can be categorized further into: (1) clustering, classification, and outlier detection; (2) association rule mining and sequential pattern mining; (3)text mining.

EDM methods that most used is web mining methods listed by Romero and Ventura are the most prominent method used in EDM research. Currently, using students' past academic performance to predict their future performance is a common task of data mining techniques (Manjarres, Sandoval, & Suárez, 2018). The aim of these tasks is to predict how well the students will do courses before registration. These predictions help students to avoid having to drop out, help the course instructor to identify students that more likely to fail and then change teaching strategies and take appropriate actions to improve student success. This research aims to provide a data-mining model for predicting student performance in advanced math, chemistry and physics courses based on their performance in English and basic mathematics courses.

Literature shows that classification and association rules mining have been applied to analyze the student data and predict students' performance. For example, depending on students' activities, the classification algorithm (the C4.5 algorithm) was used for classifying students into several groups. Meanwhile an Apriori algorithm was applied to explore association rules based on the students' grades in the course. The Apriori algorithm introduced by Agrawal et al. (1994) used to generate association rules mining. This algorithm is freely available in Weka software. The K-means algorithm was also used to cluster students from a course into different groups depending on their activities and their final marks (Bienkowski, Feng, & Means, 2012).

In (2013) Kasih, Ayub, and Susanto suggested a model to predict student(s) grade in the programming course based on the grades in the other courses. The prediction value was classified into three categories: Extraordinary, very satisfactory, and satisfactory. They Use the Apriori algorithm to explore the relationship between a programming course and other courses they took during the first four-semesters of their study period. The authors found a high correlation between programming course and math courses (Romero & Ventura, 2006).

Other researchers have proposed a model to enhance students' performances using classification association rules mining (CARM) technique. They applied the Apriori algorithm to compare the students' performance. They took a sample of student at Master of Computer Applications at the undergraduate and post-graduate levels. Using students' grades in common courses, they explored associations, and then they identified factors that determined students' chances of success or failure. They found that syllabus plan, student's interest, teaching and evaluation techniques more related to students' success (Castro, Vellido, Nebot, & Mugica, 2007). Students' demographic attributes, performance in first year courses and their overall performance using regression technique. Based on the data of 85 university students, they found a strong correlation between performance in a first-year computer science courses and the student's overall performance in the program. developed models to predict students' university performance based on students' characteristics (Romero & Ventura, 2006).

They used data of 10,330 students in the Bulgarian educational sector; students were described by 20 attributes such as gender, birth year, place of birth, place of living and university scores. Using algorithms such as decision tree C4.5, Naive Bayes, Bayesian networks, to classify the students into 5 classifications: Excellent, Very Good, Good, Average or Bad. They found that decision tree classifier performed best having the highest overall accuracy of student performance.

After using a sample of 210 undergraduate students who had enrolled in the academic batches of 2007-08 and 2008-09. They studied the performance of undergraduate students using data mining techniques such as K-means algorithm, Focusing upon two aspects of students' performance (Thakar, 2015). Predicting students' achievement at the end of a four-year study program and studying students' progressions and combining them with prediction results.

Two important groups of students have been identified: low achieving students and high achieving students. They concluded that using a small number of courses as indicators of good or poor performance will help to provide timely warnings and give support to low performance students and provide advice and opportunities to high performing students. Varghese, et al. (2010) used the k-means algorithm to cluster 8000 students based on five variables: the level of attendance, test scores, grades seminars, assignments, and students' final grades. They found a strong relationship between the levels of presence of student and their achievement grades.

Commonly Techniques Used in EDM

Various data mining methodologies, such as clustering, classification and association rule mining rules have been used to predict students' performance. These methodologies are the most used data mining techniques in EDM research (Romero & Ventura, 2007). By using these techniques, it is expected to extract important hidden information and knowledge from the data. The extracted information can be used in teaching evaluation, reduce dropout and identifying students at risk etc., thus help educators and management to establish a pedagogical basis for corrective decision making when designing or modifying a course or teaching methodology and adopt new strategies to improve student success (Anoopkumar & Rahman, 2018).

Data gathered from either a traditional classroom or a web- based educational system are two types of data that can be analyzed using data mining techniques. The following sections present overview of the data mining techniques that are most used in EDM research e.g., the K-means

algorithm for clustering, the J48 algorithm for classification, and the Apriori algorithm for association rules mining experiments.

Classification

Classification is one of the most frequently used techniques in EDM research. Classification encompasses predicting the value of an (categorical) attribute (the class) by using a model or a classifier. Training dataset can be used to build classification model, while a test dataset can be used to verify the model. Any new unlabeled class (pattern) will be labeled with a class attribute. A classification technique includes two- step process: the learning step and the classification step. Training dataset uses to build the model in the learning step, and then the model uses to predict the class label for a given dataset in the classification step (Jalota & Agrawal, 2019).

Classification has been widely used in EDM research to perform many tasks. For example: (a) classifying students into different groups with equal final marks e.g., classify or group students into three classes: high, moderate, and low; (b) exploring that the students more likely to correctly answering questions and thus success in the whole test if they read feedback for the related questions; and; (c) predicting students' exam marks by using a decision tree.

Predict academic outcomes of student performance at the end of the school year. Predicting students' performance at the university based on their personal and pre-university characteristics.

This study uses the decision tree as a popular classification technique. A decision tree algorithm consists of an algorithm that constructs a decision tree in a top-down recursive approach. This study uses the widely used C4.5 algorithm to create decision tree. C4.5 algorithm, include the superior stability between precision, speed, and interpretability of results. Quinlan invented the C4.5 algorithm since the invention of the C4.5 algorithm it has been considered as a benchmark for many newer classification algorithms. This study uses the J48 decision tree algorithm that is an Open-Source Java implementation of the C4.5 algorithm available in the Weka data-mining package. This study objective in using J48 algorithm is to: (a) build a model that can generates a classification-decision tree to be used to predict the performance of new students; (b) it is freely available in the Weka data-mining package. Weka is currently one of most popular open-source data mining systems.

Association Rules Mining

Association rules mining is also one of the most frequently techniques used by relatively large amount in EDM research. Association rules mining used to perform several tasks: Association rule mining has used to reveal relationships among the attributes in the dataset. If-then relationships use to represent these relationships among attributes and values. An association rule-mining algorithm seeks to discover a combination and/or a sequence of items that typically occur in the database.

Association rules mining has been widely used in EDM research to perform many tasks. For example: exploring association rules so that educators may use to improve the performance educational. Exploring students' mistakes that often occur together. Searching for important aspects of students' behavior when working in a group. Discover and identify all rules within some minimum support and confidence constraints. To predict student performance (mark prediction) in an e- learning environment; and to predict learner performance based on the learning portfolios compiled.

This study uses a standard Apriori algorithm as a popular classification technique. The Apriori algorithm was invented by Agrawal and Srikant (1994) and since then has become a benchmark for association rule mining research (MInfoTech, 2014). At the first step Apriori algorithm builds frequent item sets and then extract association rules from these frequent item sets. Apriori algorithm is available in the Weka data-mining package. Although, the Apriori algorithm

can generates a lot of rules in the form of IF-THEN relationships. However, not association rules mining is also one of the most frequently techniques used by relatively large amount in EDM research. Association rules mining used to perform several tasks: Association rule mining has used to reveal relationships among the attributes in the dataset. If-then relationships use to represent these relationships among attributes and values. An association rule-mining algorithm seeks to discover a combination and/or a sequence of items that typically occur in the database.

Association rules mining has been widely used in EDM research to perform many tasks. For example: exploring association rules so that educators may use to improve the performance educational. exploring students' mistakes that often occur together. searching for important aspects of students' behavior when working in a group. discover and identify all rules within some minimum support and confidence constraints. to predict student performance (mark prediction) in an e-learning environment; and to predict learner performance based on the learning portfolios compiled.

This study uses a standard Apriori algorithm as a popular classification technique. The Apriori algorithm was invented by Agrawal and Srikant (1994) and since then has become a benchmark for association rule mining research (MInfoTech, 2014). At the first step Apriori algorithm builds frequent item sets and then extract association rules from these frequent item sets. Apriori algorithm is available in the Weka data-mining package. Although, the Apriori algorithm can generates a lot of rules in the form of IF-THEN relationships. However, not all the rules generated by the algorithm are useful or interesting. Two certain requirements can be used to measure which rules are useful: support and confidence. The support (coverage) is the number of instances that the algorithm predicted correctly, while the confidence (accuracy) is the number of instances that the algorithm predicted correctly as the proportion of the whole instances of the predicted attribute in this study, these two measurements are used to select the rules that were considered interesting and useful.

CONCLUSION

This Research have reviewed EDM definitions, purpose of EDM research, different applications of EDM and the prominent relevant existing works in the EDM research. This review provides guidelines and theoretical foundation for this study. It identified the positive aspects and pedagogical aspects of the previously research that can be followed in this study. The research concluded with a discussion of the common methods that used in EDM research that can be followed in this study.

ACKNOWLEDGMENT

REFERENCES

Ahmed, A. M., Rizaner, A., & Ulusoy, A. H. (2016). Using data mining to predict instructor performance. Procedia Computer Science, 102, 137-142.

Angeli, C., Howard, S. K., Ma, J., Yang, J., & Kirschner, P. A. (2017). Data mining in educational technology classroom research: Can it make a contribution? Computers & Education, 113, 226-242.

Anoopkumar, M., & Rahman, A. (2018). Model of Tuned J48 Classification and Analysis of Performance Prediction in Educational Data Mining. Int. J. Appl. Eng. Res, 13(20), 14717-14727.

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers & Education, 113, 177-194.

Badr, G., Algobail, A., Almutairi, H., & Almutery, M. (2016). Predicting students' performance in university courses: a case study and tool in KSU mathematics department. Procedia Computer Science, 82, 80-89.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief. Office of Educational Technology, US Department of Education.

Burgos, C., Campanario, M. L., de la Peña, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. Computers & Electrical Engineering, 66, 541-556.

Castro, F., Vellido, A., Nebot, A., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. In Evolution of teaching and learning paradigms in intelligent environment (pp. 183-221): Springer.

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior, 73, 247-256.

da Fonseca Silveira, R., Holanda, M., de Carvalho Victorino, M., & Ladeira, M. (2019). Educational data mining: Analysis of drop out of engineering majors at the UnB-Brazil. Paper presented at the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA).

Deeva, G., & De Weerdt, J. (2018). Understanding automated feedback in learning processes by mining local patterns. Paper presented at the International Conference on Business Process Management.

Dubey, R. (2016). PERFORMANCE EVALUATION of MILITARY TRAINING EXERCISES USING DATA MINING. In.

Jalota, C., & Agrawal, R. (2019). Analysis of educational data mining using classification. Paper presented at the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon).

Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. Computers in Human Behavior, 92, 496-506.

Khasanah, A. U. (2017). A comparative study to predict student's performance using educational data mining techniques. Paper presented at the IOP Conference Series: Materials Science and Engineering.

Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. (2018). Data mining techniques applied in educational environments: Literature review. Digital Education Review(33), 235-266.

Menon, A., & Islam, N. (2017). An investigation of the relationship between online activity on Studi. se and academic grades of newly arrived immigrant students: An application of educational data mining. In.

Nebot, À., Mugica, F., & Castro, F. (2020). An e-Learning Toolbox Based on Rule-Based Fuzzy Approaches. Applied Sciences, 10(19), 6804.

Peral Cortés, J., Maté, A., & Marco Such, M. (2017). Application of Data Mining techniques to identify relevant Key Performance Indicators.

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. F. (2020). Investigating variation in learning processes in a FutureLearn MOOC. Journal of computing in higher education, 32(1), 162-181.

Romero, C., & Ventura, S. (2006). Data mining in e-learning (Vol. 4): WIT press.

Thakar, P. (2015). Performance analysis and prediction in educational data mining: A research travelogue. arXiv preprint arXiv:1509.05176.

Wang, R., Ji, W., Liu, M., Wang, X., Weng, J., Deng, S., . . . Yuan, C.-a. (2018). Review on mining data from multiple data sources. Pattern recognition letters, 109, 120-128.

Yang, F., & Li, F. W. (2018). Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. Computers & Education, 123, 97-108.