

Educational Data Mining Applications and Techniques

Fatima Alshareef¹, Hosam Alhakami², Tahani Alsubait³, Abdullah Baz⁴
College of Computer and Information Systems
Umm Al-Qura University, Makkah, Saudi Arabia

Abstract—Educational data mining (EDM) uses data mining techniques to analyze huge amounts of student data in the educational environments. The main purpose of EDM is to analyze and solve educational issues and, consequently, improve educational processes. With the emergence of EDM applications in the educational environments, several techniques have been identified to implement these applications. This paper reviews the relevant studies in EDM including datasets and techniques used in those studies and identifies the most effective techniques. The most prevalent applications include predicting student performance, detecting undesirable student behaviors, grouping students and student modeling. These applications aim to help decision makers in the educational institutions to understand student situations, improve students' performance, identify learning priorities for different groups of students and develop learning process. The prediction accuracy is selected as the evaluation criteria for the effectiveness of educational data mining techniques. The results show that Bayesian Network and Random Forest are the most effective techniques for predicting student performance, Social Network Analysis is the best technique for detecting undesirable student behaviors, Clustering and Social Network Analysis are the most effective techniques for grouping students and student modeling, respectively. This study recommends conducting more comprehensive and extended studies to evaluate the effectiveness of EDM techniques with an extended evaluation criteria.

Keywords—Educational data mining; student performance; prediction; classification; clustering

I. INTRODUCTION

The main aim of educational systems is providing knowledge and skills for students to move into their future careers in a specific period. The way that the educational systems meet effectively with this aim is a key determinant for both social and economic progress [1]. The used technologies in educational systems generated massive data that is difficult to analyze with a human eye [2]. Educational data mining (EDM) uses different data mining techniques [3] to analyze students' data in educational environments. The main purpose of EDM is to analyze and solve educational issues and, consequently, improve educational processes [4]. Therefore, its goal is to examine the educational data for resolving the associated issues with education. EDM include extracting useful, interpretable, interesting and novel information from data within the educational field.

Lately, with the emergence of the educational data mining applications in educational environments, several techniques have been identified to implement these applications. Therefore, this study explores the EDM techniques of educational data mining applications. Among these applications, the most

prevalent are predicting student performance, detecting undesirable student behaviors, grouping students and student modeling.

There are different applications in EDM where these applications have different objectives including enhancing and improving the quality of learning and improve the learning process understanding process [5]. Moreover, EDM applications target the different stakeholders in the educational systems including students, researchers, administrators and educators. Providing recommendations, personalization and feedback can develop the students learning.

This paper reviews the relevant studies in the EDM landscape including the datasets and techniques used in those studies, and identifies the most effective techniques for educational data mining applications, with an emphasis on applications concerning students. The importance of these applications is that they help decision makers in educational institutions to gain a deep understanding of student situations, improve students' performance, identify learning priorities for different groups of students and develop learning process.

This paper is organized as follows. Section II presents a review of some studies that analysed the landscape of Educational Data Mining (EDM). Section III is devoted for presenting a thorough analysis of EDM applications and associated techniques. Section IV discusses surveyed techniques for each application and identifies the most effective techniques, while Section V concludes this research considering some thoughts for future work.

II. RELATED WORK

Several investigations have been carried out regarding Educational Data Mining (EDM) applications and techniques in academic environments, demonstrating the importance of EDM for extracting accurate information about students' behavior and effectiveness of the learning process [6]. There are also several survey papers published about EDM so far. Recently in one of the surveys published in 2018 by Ray and Saeed [2] on EDM in higher education, four areas of applications have been counted, namely predicting students' performance, educating students using big data, assessment of students' learning, and teaching and research. Based on these applications, they described EDM techniques that have been applied in order to improve and understand the learning process of students. In another survey about EDM applications written by Bakhshinategh et al. [5], 13 categories of EDM applications were suggested under student modeling, decision support systems and other applications. This survey has been

useful because it described these applications with the help of research examples. Due to the importance of educational data mining, many researches related to education involved analysis and data mining. This paper reviews EDM techniques with an emphasis on their applications concerning students.

III. EDUCATIONAL DATA MINING APPLICATIONS AND ASSOCIATED TECHNIQUES

A. Predicting Student Performance

The main aim of this application is predicting the students' academic failure to improve their learning and develop the educational process. Also, it helps stakeholders in education to improve the performance of students in future [7] [5]. The most used techniques for predicting student performance includes Bayesian classification, decision trees, neural network, rule based and feature selection.

In 2015, Ahmad, Ismail and Abdul Aziz [8] studied the techniques for predicting students' academic performance in the first year of the bachelor's in computer science. They used different techniques including Naive Bayes, Rule Based, and Decision Tree for applying them on the data of students to produce the best prediction model for students' academic performance. The results of this study showed that the Rule Based is the best prediction model as it received the highest percent of prediction accuracy 71.3%. The study [9] by Kaur, Singh and Josan focused on predicting slow learners among students by different classification techniques. They collected a dataset of 152 students from a high school then they analyzed and tested students' performance using WEKA tool. The result of the comparison between used predictive techniques showed that Multilayer Perception technique had the best prediction accuracy of 75%.

The study [10] by Mueenm Zafar and Manzoor in 2016 applied different techniques of data mining including Naive Bayes, Decision Tree and Neural Network on students' data for two courses to predict the academic performance of students. The results showed that Naive Bayes algorithm had the highest prediction accuracy of 86%, while Decision Tree and Neural Network had 82.7% and 79.2% respectively.

In 2017, Abu Amra and Maghari in their study [11] proposed the best model for predicting students' performance based on their attributes by using Naive Bayes and K -Nearest Neighbor (KNN) classification algorithms. They collected data from 500 students with eight attributes in secondary school. The result of their study showed that Naive Bayes classifier had better prediction accuracy of 93.6% while KNN classifier had 62.9%. Another study [12] by Almarabeh applied five data mining classification techniques for predicting students' performance using WEKA tool. He collected data from 225 students in university. The result of his study showed that Bayesian Network had the highest prediction accuracy of 92%, Naive Bayes and J48 had the same prediction accuracy of 91.11%, Neural Network had the prediction accuracy of 90.2% and Iterative Dichotomiser 3 (ID3) had the the lowest prediction accuracy of 88%. Mousa and Maghari in their study [13] applied three data mining classification techniques and they are Naive Bayes, Decision Tree and K -Nearest Neighbor (K-NN) to predict the performance of students using academic attributes. They collected data from a preparatory in Gaza

strip school of 1100 male students. Their result showed that Decision Tree had the best prediction accuracy of 92.96%, Naive Bayes had the prediction accuracy of 91.50% and K-NN had the prediction accuracy of 90.91%. Kapur and Ahluwalia [14] compared six data mining algorithms to predicting the marks of students and they are Naive Bayes, Random Forest, Decision Tree, IBk, K-star and Naive Bayes Multiple Nominal. They collected a dataset of 480 records with 16 attributes, and they used WEKA tool. The results of their study showed that Random Forest had the highest prediction accuracy of 76.667%. Also, Khasanah and Harwati [15] applied Bayesian Network and Decision Tree techniques for predicting the performance of students to avoid students failures. They collected data from Industrial Engineering students at Islam Indonesia University. The result showed that Bayesian Network had the best prediction accuracy of 98.08% while Decision Tree had 94.23%. Makhtar, Nawang and Shamsuddin in their study [16] classified students' performance according to their performance in specific subjects using Naive Bayes algorithm. They collected 488 student's data from the Maktab Rendah Sains MARA Kuala Berang Information System. Their result showed that Naive Bayes method had accuracy of 73.4%.

Hussain, Dahan, Ba-Alwib and Ribata in 2018 [17] applied four data mining techniques to predict the students' performance and avoid students' dropout using WEKA tool. They collected a dataset of 300 students and 24 attributes at India and Assam colleges. Their results showed that Random Forest had the highest prediction accuracy of 99%, Bayes Network had prediction accuracy of 65.33%, J48 had prediction accuracy of 73% and PART had prediction accuracy of 74.33%.

In 2019, Salal, Abdullaev and Kumar [18] implemented data mining classification algorithms including Naive Bayes, Random Forest, JRip, REPTree, OneR, Decision Tree (J48), SimpleLogistic and ZeroR for predicting students' academic performance. They collected 649 student's data with 33 attributes from two secondary school then they analyzed it using WEKA tool. The result showed that Decision Tree (J48), REPTree and OneR had the prediction accuracy of more than 76%. Decision Tree (J48) had accuracy of 76.2712%, REPTree and OneR had the same accuracy of 76.7334%. The study [19] by Agarwal, Maheshwari, Roy, Pandey and Rautray analyzed 306 students' data in higher education for predicting student performance using two classification algorithms K-Nearest Neighbor and Random Forest. The result of their study showed that Random Forest had the highest prediction accuracy of 93.54%. Adekitan and Salau [20] analyzed the performance of students using six data mining algorithms like Naive Bayes, Random Forest, Logistic Regression, Decision Tree, Neural Network and Tree Ensemble. They collected 1841 data from engineering students in the first three years. They result showed that Logistic Regression had the maximum prediction accuracy of 89.15%. Adekitan and Noma-Osaghae [21] used data mining algorithms for predicting student's performance in KNIME and Orange platforms. They analyzed student's data in their first year at Covenant University in Nigeria. The result of their study showed that Logistic Regression in KNIME platform and Neural Network in Orange platform had the prediction accuracy of 50.23% and 51.9% respectively. Another study [22] by Rifat, Al Imran and Badrudduza used six classification algorithms of data mining including Random Forest, Decision Tree, Tree Ensemble, Gradient Boosted Tree,

K-Nearest Neighbors and Support Vector Machine for predicting the students' performance. They collected 398 business students' data from the Marketing department of a renowned university in Bangladesh then they analyzed it using KNIME, the Konstanz tools. The result of their study showed that Random Forest had the highest prediction accuracy of 94.1%.

In 2020, Alhakami et al. in their study [23] used J48, Naive Bayes algorithms to predict students' academic performance and help in advising students using WEKA tool. They collected 38671 students' data of both male and female from Umm Al-Qura University for 5 years, with several attributes including Exams Marks, School, Sex, Age, Nationality, City and final grade. Their result showed that J48 algorithm had the highest accuracy of 84.38%.

B. Detecting Undesirable Student Behaviors

With its similarity to student performance prediction, this application focuses on detecting undesirable student behaviors including erroneous actions, low motivation, cheating, academic failure. The main used data mining techniques including classification, clustering, outlier detection and feature selection, decision tree and neural networks [5].

In 2012, Bayer et al. focused in their study [24] on predicting school failures and drop-outs when enriching student data with derived data from student's social behavior. The collected data described gathered social dependencies from discussion board conversations and e-mail mainly. They described new features extraction from both represented behavior and data student data through a constructed social graph. Then, a novel method was introduced for learning a student failure prediction classifier that uses cost-sensitive learning in order to lower the wrongly classified unsuccessful students number. The results showed that Social Network Analysis (SNA) produced significant increase in the prediction accuracy to 92.89%.

The study [25] by Guarín, Guzmán and González in 2015 applied two data mining methods including Naive Bayes and decision Tree to predicting low academic performance of students in the first four enrollments. They collected data from Architectural Engineering program and Computer and System Engineering program. The result of their study showed that Naive Bayes had the best prediction accuracy of 75%.

The study [26] by Athani et al. in 2017 aimed to enhance the behavior of secondary school students using techniques of data mining. Naive Bayesian classifier is implemented to predict the behavior of students to create the prediction. The classifier accuracy is calculated through WEKA tool where confusion matrix was generated. The obtained classifier accuracy is 87% which could be further enhanced through appropriate attributes selection.

In 2019, the study [27] by Pattanaphanchai, Leelertpanyakul and Theppalak proposed a model to predict students' dropout patterns using WEKA tool. The dataset is collected from Faculty of Science, Prince of Songkla University of five years. The result of their study showed that JRip had a prediction accuracy of 77.30%.

C. Grouping Students

The aim of grouping students application is to create a group or cluster of students according to different profile in-

formation properties [28]. This application is used by different stakeholders in education for several tasks to develop the educational process [29]. It is often unlike clustering similar students together where the aim is to group students who complement each other. Moreover, the highest dissimilarity is detected between clusters when clustering students. However, this is not always the case in students grouping [5]. The most common EDM techniques used in grouping students include clustering, neural network and feature selection.

In 2013, Harley, Trevors and Azevedo in their study [30] presented the obtained results using clustering on collected data from 106 students. Three extracted clusters were analyzed and validated through multivariate statistics (MANOVAs) to characterize the three distinct students profiles, showing statistically significant differences of all the twelve used variables for the formation of the clusters (such as: performance, note-taking use and sub-goals attempted number). The results showed that variations existed among the clusters concerning perceived prompts through the system. The prediction accuracy of the overall clusters was 78.8%. On the other hand, Nunes and Minussi in their study [31] investigated a Neural Network technique used in students grouping who are likely to be failed. The results of this study indicated that the Neural Network technique reached prediction accuracy of 76%.

D. Student Modeling

This application defines different aspects characterizing the student including cognition, skills, emotions, domain knowledge, learning strategies, achievements, features, learning preferences, affects and evaluation. The aim is to characterize the student and adjust the teaching processes to meet the learning requirements of students [32]. The main used techniques in students modeling application including Social Network Analysis (SNA), rules induction, decision tree, Linear Discriminant Analysis (LDA), Bayes theorem and rules induction.

Dekker, Pechenizkiy and Vleeshouwers (2009) in their study [33] aimed to predict the students drop using classification algorithms of data mining. They collected dataset of 648 students in the Electrical Engineering program. The results of this study showed that simple decision trees and intuitive decision trees classifier gave prediction accuracy of 75% and 80% respectively.

In another study [34] by Macfadyen and Dawson (2010), the researchers included an investigation of which student's prediction modeling is more effective. The results of the study showed that Social Network Analysis (SNA) has generated the best prediction. Logistic modeling validated Social Network Analysis (SNA) predictive power that achieved prediction accuracy of 81%.

The study [35] by Sivakumar, Venkataraman and Selvaraj (2016) used improved decision tree for predictive modeling of dropout students. The dataset is collected from 240 records at university in India. The results showed that improved decision had a prediction accuracy of 97.50%.

TABLE I. COMPARISON BETWEEN THE PREDICTION ACCURACY OF THE EDM TECHNIQUES.

EDM Applications	Techniques	The highest Prediction Accuracy	References
Predicting student performance	Naïve Bayes	93.6%	[11]
	Bayesian Network	98.08%	[15]
	Decision Tree	94.23%	[15]
	Rule Based	71.3%	[8]
	Neural Network	90.2%	[12]
	K -Nearest Neighbor(KNN)	90.91%	[13]
	Multilayer Perception	75%	[9]
	REPTree	76.7334%	[18]
	OneR	76.7334%	[18]
	Iterative Dichotomiser 3 (ID3)	88%	[12]
	Random Forest	99%	[17]
	PART	74.33%	[17]
	Logistic Regression	89.15%	[20]
	Detecting undesirable student behaviors	Social Network Analysis (SNA)	92.89%
Naïve Bayesian		87%	[26]
JRip		77.30%	[27]
Grouping students	Clustering	78.8%	[30]
	Neural Network	76%	[31]
Student modeling	Social Network Analysis (SNA)	81%	[34]
	Decision Trees	80%	[33]

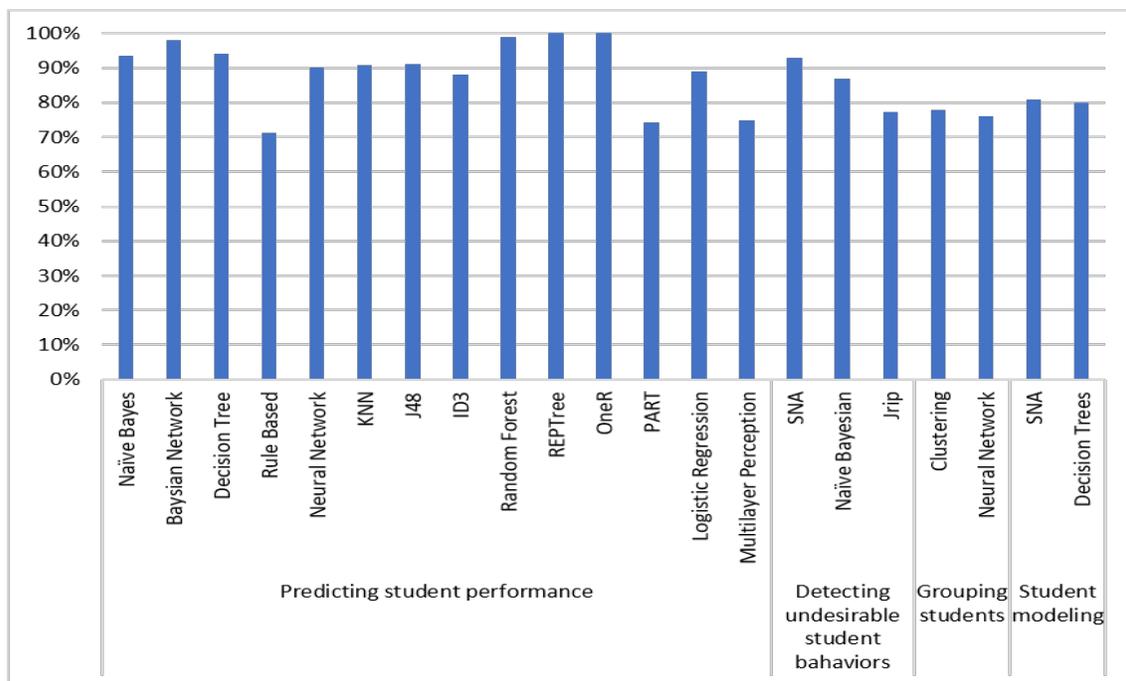


Fig. 1. Prediction Accuracy of the EDM techniques.

IV. DISCUSSION

Based on the above review of studies which analyzed student’s data to solve some educational issues, seven techniques are identified for four EDM applications and they include:

- Predicting Student Performance
- Detecting Undesirable Student Behaviors
- Grouping Students
- Student Modeling

The comparison between the surveyed techniques was based on the prediction accuracy as a common evaluation criteria in each of the selected research papers. “Table I” demonstrates the main surveyed techniques.

According to “Fig. 1”, Social Network Analysis (SNA) is the most common technique where it is the most effective technique used in detecting the undesirable student behaviors as well as student modeling. However, the efficiency of Social Network Analysis (SNA) is not the same for both applications, which indicates that each application implies different type of techniques. Social network analysis (SNA) focuses on

detecting the students' interaction pattern and has been evolved increasingly with the emergence of social networking such as Facebook and Twitter [36] [37]. Also, SNA is used for assessing the student's participation in discussions of online courses [38]. In the previous studies, SNA was used to monitor students' creative capacity [39], to detect "at risk" students besides Bayer et al. [24] who used SNA for dropout's prediction.

Naive Bayes is a classification algorithm of educational data mining techniques. Naive Bayes algorithm is mainly used for predictive modeling that is based on the Bayesian techniques with independent attributes [40]. Naive Bayes, Rule induction and decision tree algorithms can easily be implemented in the IF-THEN rules form of object-oriented programming that can be simply understood [41] [42]. This way, normal users who have no deep knowledge regarding data mining can understand easily the obtained results using the previous algorithms.

Decision tree is a popular and powerful prediction and classification technique. It is the most frequent data mining technique that is used in the related studies [43]. This algorithm consists of number of nodes which are used to getting related information for supports the decision-making process on the root node [44].

Neural Network is a set of input/output units that are connected where every connection has a different associated weight with each other [45]. Throughout the learning part, the network learns through modifying the weights in order to predict the input samples correct class.

Clustering techniques achieved better prediction accuracy compared to neural network techniques in students' grouping application. Clustering is a very efficient technique in grouping where it divides the data into groups or clusters with similar characteristics [46]. However, the size of data is reduced when clustering [47], so some details are lost.

The results of this review paper could be a reference for decision makers in educational systems, where such data could provide a decision support validation for the used prediction technique. Moreover, this paper analyses the efficiency of the educational data mining applications with the right predictions.

This study helps in improving and enhancing the performance and experience of students by presenting best prediction techniques. Teachers also can benefit from this paper by finding the best methods that they can use to develop the educational processes. Using educational data mining, Teachers can identify the students' behavioral patterns that can support their judgments and teaching methods and determine indicators of student engagement and satisfaction besides monitoring the learning progress. Moreover, researchers can use this review paper for further concentration on the evaluation and development of educational data mining techniques.

V. CONCLUSION AND FUTURE WORK

Recently, with the increase of utilizing data mining applications in educational environments, this paper identifies the most effective techniques for each of these EDM applications. The importance of this review paper lies on the united evaluation criteria for the comparison of the different techniques

for each EDM application. Prediction accuracy is used as an indicator for the effectiveness of the surveyed techniques. This paper indicates that the effective technique in one application does not necessarily means it will be effective on another application. Therefore, further surveys should be conducted for each of the EDM applications to more accurately identify the most effective techniques. Moreover, extended evaluation and comparison criteria should be used in the evaluation.

REFERENCES

- [1] A. Algarni, "Data mining in education," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 6, pp. 456–461, 2016.
- [2] S. Ray and M. Saeed, "Applications of educational data mining and learning analytics tools in handling big data in higher education," in *Applications of Big Data Analytics*, pp. 135–160, Springer, 2018.
- [3] A. V. Manjarres, L. G. M. Sandoval, and M. S. Suárez, "Data mining techniques applied in educational environments: Literature review," *Digital Education Review*, pp. 235–266, 2018.
- [4] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, 2010.
- [5] B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," *Education and Information Technologies*, vol. 23, no. 1, pp. 537–553, 2018.
- [6] I. F. Siddiqui, Q. A. Arain, et al., "Analyzing students' academic performance through educational data mining.," *3C Tecnologia*, 2019.
- [7] K. Karthikeyan and K. PALANIAPPAN, "On improving student performance prediction in education systems using enhanced data mining techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 5, 2017.
- [8] F. Ahmad, N. H. Ismail, and A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques," *Applied Mathematical Sciences*, vol. 9, no. 129, pp. 6415–6426, 2015.
- [9] P. Kaur, M. Singh, and G. S. Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector," *Procedia Computer Science*, vol. 57, pp. 500–508, 2015.
- [10] A. Mueen, B. Zafar, and U. Manzoor, "Modeling and predicting students' academic performance using data mining techniques," *International Journal of Modern Education and Computer Science*, vol. 8, no. 11, p. 36, 2016.
- [11] I. A. A. Amra and A. Y. Maghari, "Students performance prediction using knn and naïve bayesian," in *2017 8th International Conference on Information Technology (ICIT)*, pp. 909–913, IEEE, 2017.
- [12] H. Almarabeh, "Analysis of students' performance by using different data mining classifiers," *International Journal of Modern Education and Computer Science*, vol. 9, no. 8, p. 9, 2017.
- [13] H. Mousa and A. Y. Maghari, "School students' performance prediction using data mining classification," *School Students' Performance Prediction Using Data Mining Classification*, vol. 6, no. 8, 2017.
- [14] B. Kapur, N. Ahluwalia, and R. Sathyaraj, "Comparative study on marks prediction using data mining and classification algorithms," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 3, 2017.
- [15] A. U. Khasanah et al., "A comparative study to predict student's performance using educational data mining techniques," in *IOP Conference Series: Materials Science and Engineering*, vol. 215, p. 012036, IOP Publishing, 2017.
- [16] M. MAKHTAR, H. NAWANG, and S. N. WAN SHAMSUDDIN, "Analysis on students performance using naïve bayes classifier," *Journal of Theoretical & Applied Information Technology*, vol. 95, no. 16, 2017.
- [17] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using weka," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 2, pp. 447–459, 2018.

- [18] Y. Salal, S. Abdullaev, and M. Kumar, "Educational data mining: Student performance prediction in academic," *IJ of Engineering and Advanced Tech*, vol. 8, no. 4C, pp. 54–59, 2019.
- [19] K. Agarwal, E. Maheshwari, C. Roy, M. Pandey, and S. S. Rautray, "Analyzing student performance in engineering placement using data mining," in *Proceedings of International Conference on Computational Intelligence and Data Engineering*, pp. 171–181, Springer, 2019.
- [20] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," *Heliyon*, vol. 5, no. 2, p. e01250, 2019.
- [21] A. I. Adekitan and E. Noma-Osaghae, "Data mining approach to predicting the performance of first year student in a university using the admission requirements," *Education and Information Technologies*, vol. 24, no. 2, pp. 1527–1543, 2019.
- [22] M. R. I. Rifat, A. Al Imran, and A. Badrudduza, "Educational performance analytics of undergraduate business students," *International Journal of Modern Education and Computer Science*, vol. 11, no. 7, p. 44, 2019.
- [23] H. Alhakami, T. Alsubait, and A. Aljarallah, "Data mining for student advising," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 2020.
- [24] J. Bayer, H. Bydzovská, J. Géryk, T. Obsivac, and L. Popelinsky, "Predicting drop-out from social behaviour of students.," *International Educational Data Mining Society*, 2012.
- [25] C. E. L. Guarín, E. L. Guzmán, and F. A. González, "A model to predict low academic performance at a specific enrollment using data mining," *IEEE Revista Iberoamericana de tecnologías del Aprendizaje*, vol. 10, no. 3, pp. 119–125, 2015.
- [26] S. S. Athani, S. A. Kodli, M. N. Banavasi, and P. S. Hiremath, "Student academic performance and social behavior predictor using data mining techniques," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 170–174, IEEE, 2017.
- [27] J. Pattanaphanchai, K. Leelertpanyakul, and N. Theppalak, "The investigation of student dropout prediction model in thai higher education using educational data mining: A case study of faculty of science, prince of songkla uni-versity," *Journal of University of Babylon for Pure and Applied Sciences*, vol. 27, no. 1, pp. 356–367, 2019.
- [28] D. Krpan and S. Stankov, "Educational data mining for grouping students in e-learning system," in *Proceedings of the Iui 2012 34th International Conference on Information Technology Interfaces*, pp. 207–212, IEEE, 2012.
- [29] P. Nithya, B. Umamaheswari, and A. Umadevi, "A survey on educational data mining in field of education," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 5, no. 1, pp. 69–78, 2016.
- [30] J. M. Harley, G. J. Trevors, R. Azevedo, *et al.*, "Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning," *JEDM— Journal of Educational Data Mining*, vol. 5, no. 1, pp. 104–146, 2013.
- [31] V. R. Martinho, C. Nunes, and C. R. Minussi, "Prediction of school dropout risk group using neural network," in *2013 Federated Conference on Computer Science and Information Systems*, pp. 111–114, IEEE, 2013.
- [32] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert systems with applications*, vol. 41, no. 4, pp. 1432–1462, 2014.
- [33] G. W. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers, "Predicting students drop out: A case study.," *International Working Group on Educational Data Mining*, 2009.
- [34] L. P. Macfadyen and S. Dawson, "Mining lms data to develop an "early warning system" for educators: A proof of concept," *Computers & education*, vol. 54, no. 2, pp. 588–599, 2010.
- [35] S. Sivakumar, S. Venkataraman, and R. Selvaraj, "Predictive modeling of student dropout indicators in educational data mining using improved decision tree," *Indian Journal of Science and Technology*, vol. 9, no. 4, pp. 1–5, 2016.
- [36] D. Garcia-Saiz, C. Palazuelos, and M. Zorrilla, "Data mining and social network analysis in the educational field: An application for non-expert users," in *Educational data mining*, pp. 411–439, Springer, 2014.
- [37] M. Saqr and A. Alamro, "The role of social network analysis as a learning analytics tool in online problem based learning," *BMC medical education*, vol. 19, no. 1, p. 160, 2019.
- [38] S. Slater, S. Joksimović, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for educational data mining: A review," *Journal of Educational and Behavioral Statistics*, vol. 42, no. 1, pp. 85–106, 2017.
- [39] S. Dawson, J. P. L. Tan, and E. McWilliam, "Measuring creative potential: Using social network analysis to monitor a learners' creative capacity," *Australasian Journal of Educational Technology*, vol. 27, no. 6, 2011.
- [40] A. Hamoud, A. Humadi, W. A. Awadh, and A. S. Hashim, "Students' success prediction based on bayes algorithms," *Hamoud, AK, et al., Students' Success Prediction based on Bayes Algorithms. International Journal of Computer Applications*, vol. 178, no. 7, pp. 6–12, 2017.
- [41] L. P. Khobragade and P. Mahadik, "Students' academic failure prediction using data mining," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 11, pp. 290–298, 2015.
- [42] V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced naive bayes model," in *International Conference on Intelligent Data Engineering and Automated Learning*, pp. 194–201, Springer, 2013.
- [43] T. Jeevalatha, N. Ananthi, and D. S. Kumar, "Performance analysis of undergraduate students placement selection using decision tree algorithms," *International Journal of Computer Applications*, vol. 108, no. 15, 2014.
- [44] M. Kavitha, D. Raj, *et al.*, "Educational data mining and learning analytics-educational assistance for teaching and learning," *arXiv preprint arXiv:1706.03327*, 2017.
- [45] A. D. Kumar, R. P. Selvam, and K. S. Kumar, "Review on prediction algorithms in educational data mining," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 8, pp. 531–537, 2018.
- [46] R. Saxena, "Educational data mining: Performance evaluation of decision tree and clustering techniques using weka platform," *International Journal of Computer Science and Business Informatics*, 2015.
- [47] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *IEEE Access*, vol. 5, pp. 15991–16005, 2017.