Journal of Engineering and Sustainable Development
https://jeasd.uomustansiriyah.edu.iq/index.php/jeasd
https://doi.org/10.31272/jeasd.26.5.11

Vol. 26, No. 05, September 2022
ISSN 2520-0917

A PREDICITON MODEL BASED ON STUDENTS'S BEHAVIOR IN E-LEARNING ENVIRONMENTS USING DATA MINING TECHNIQUES

*** Anwar Adnan Alnawas[1] Mohammed M H Al-Jawad[2] Hasanein Alharbi[3]**

1) Nasiriyah Technical Institute, Southern Technical University, Iraq
2) University of Kerbala, College of Computer Science and Information Technology, Karbala, 56001, Iraq
3) University of Babylon, IT College, Babylon, 51002, Iraq

Received 16/5/2022 ***Accepted in revised form 23/5/2022*** ***Published 1/9/2022***

Abstract: E-Learning has become an essential teaching approach during COVID-19 pandemic. All over the world, various internet-based learning management systems (Google classroom, Moodle, etc.) were adopted to convey knowledge and enhance learning outcomes. However, measuring learning outcomes and knowledge acquisition in E-Learning environment is a controversial issue. To this end, this paper aims to predict learning outcomes using data mining techniques. Student data are collected and analyzed to construct the prediction model. The collected data covered students from various undergraduate studies. Cross-Industry Standard Process for Data Mining is used as a research model. The obtained result shows the significant of some attributes in predicting learning outcomes. Four correlation-based attributes selection schemas are applied. The selected attributes are examined using four data mining algorithms: random forest, k-nearest neighbors, Decision Tree and neural network. The overall performance of the constructed mining models is evaluated using various performance measures: Accuracy, Precision, Recall and F1-score are calculated. Overall, an 86% accuracy is secured.

Keywords: *Data Mining; e-learning; prediction; user behavior*

1. Introduction

Covid-19 pandemic affected societies over the world in all life aspects. Higher Education sector has been significantly affected by the pandemic [1,2]. For instance, face-to-face learning approach became banned and restricted in many countries. Hence, E-Learning played an important role to overcome the rising challenges [3].

E-Learning has been increased by 53% during the pandemic [4]. It is the process of acquiring knowledge using ICT (Information and Communication Technology) infrastructures. ICT aids to make the learning content accessible via network such as Internet, Intranet, Extranet, V-sat and others [5].

E-learning has numerous advantages. It supports remote communications between teachers and students, both synchronous and asynchronous communications can be implemented. Blending learning is also achievable. Cost, administrative tasks, time and geographical constraints can all be reduced using E-Learning concept [4, 6] .

Despite its significant advantages, E-learning still has some serious challenges in developing countries. Poor ICT infrastructures, Experience in using ICT, E-readiness to adopt E-Learning pedagogical method, maintain a consistent communication with students, practical and

rresponding Author: anwar.alnawas@stu.edu.iq

clinical trainings and students' assessment are examples of E-Learning challenges [6, 7]. In fact, measuring learning outcomes and knowledge acquisitions is a fundamental challenge in E-Learning environment, [8] stated that measuring learning outcomes is one of the online assessment challenges .

Avcı et al. [9] illustrated the content of the engagement of students by three factors; emotional, behavioral and cognitive engagement. The involvement in activities, participating and observed behavior are related to behavioral engagement. That engagement and participating which are represented the student behavior will affect the online learning outcomes.

Recent research Kumar & Chong[10], Xiao et al [11] suggested that Data Mining (DM) techniques are an effective approach in analyzing educational data and predicting learning outcomes. Hence, this paper applied various DM algorithms to predict the performance of students in Iraqi undergraduate educational institutes during COVID 19 pandemic.

Orange machine learning and data mining suite Demšar et al.[12] is used to execute RF, KNN, DT and NN mining algorithms. The obtained accuracy, 86%, indicates a promising finding. The main contribution of this paper is to find the correlated parameters to help academic institutions to enhance students' outcomes.

2. Related Work

A large and growing body of literature has investigated the role of data mining in predicting learning outcomes in higher education. In this section the most relevant works is studied. For instance, D. F. Murad et al. [13] used the predicted result to provide recommendations to students by using User-Item Collaborative Filtering System which can enhance the learning out comes.

In slightly different approach, the authors of Mai et al. [14], Zeineddine et al. [15] predicted the learning outcomes using two different data mining techniques. However, the obtained accuracy, 79% and 75.9%, are moderate.

Similarly, S. Chayanukro et al. [16] analyzed Moodle logs data. The collected data are investigated using six data mining algorithms. Yet, the obtained accuracy is less than 50%.

Blending learning has been studied as well, Y. Luo et al. [17] utilized the LMS data and the administrative system to predict learning outcome. Likewise, the authors of [18] executed random forests algorithm to anticipate whether students will obtain a bachelor's degree. Eventually, a 78.84% accuracy was obtained.

P. Dabhade et al. [19] argued that educational data mining technique can be used to enhance academic education. Accordingly, they developed a data mining model to monitor students' performance. Their ultimate aim is to identify students who do not meet the expectations so a special care can be offered to improve students' performance. Multiple linear regression and support vector regression algorithms are executed. The obtained results show that there is a sold relation between student's behavior and academic performance. The linear support vector regression algorithm secured the best accuracy of 83.44%.

Alternatively, predicting students' performance before the start of the course have been investigated by A. Khan et al. [20]. The previous backlog, estimated teaching quality, ease of scoring, student quality and domain knowledge are utilized to construct the classification model. The model is called random wheel. The proposed model achieved an overall accuracy of 66%.

Furthermore, the authors of Hussain et al.[21] implemented various Machine Learning (ML)

classification and grouping techniques. Moodle data are used to detect low performing students prior to the tests. The experimental result showed that Fuzzy Unordered Rule Induction Algorithm (FURIA) ranking technique achieved the highest accuracy. Various student's categories predicted as well.

All in all, a lot of effort spent in predicting student's learning outcomes. So, educational institutes can take the required actions to enhance the learning quality. The reviewed literature showed that, P. Dabhade et al. [19] achieve the best accuracy of 83.44%

3. Methodology

CRISP-DM methodology is followed, which enhance the performance of the projects of DM by making them manageable, repeatable, less expensive and faster [22]. CRISP-DM divided to six phases, Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. Orange DM tool also used to implement this study, which facilitate the deployment and evaluation of machine learning algorithms by using a multitude of operators to prepare the datasets and import them Thange et al.[23], as shows in Figure 1 [24].

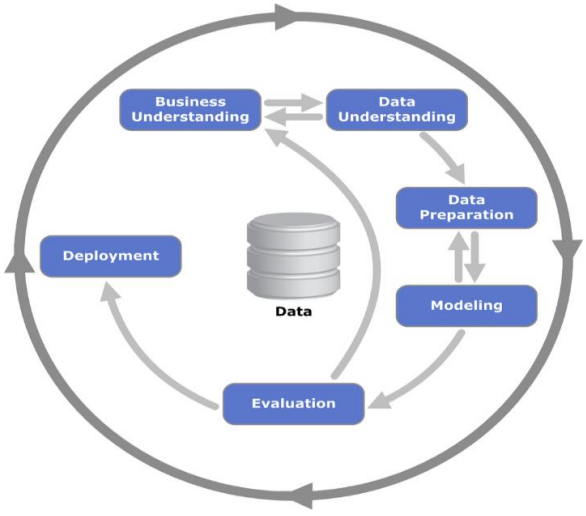

Figure 1. CRISP-DM Process Diagram

3.1. Business Understanding

This fundamental phase concentrating on satisfying the project objectives and requirements from business point of view, and the squired knowledge transformed into a well-defined of Data Mining problem by preparing an initial plan of the project to reach the outlined of the objectives [25].

Thus, the main goal of the conducted study is to predict whether students have pass or fail the final exam, where the target attribute "result" can be "0" or "1", where "0" means failed in the final exam and "1" means pass the final exam. This prediction fits in the scope of binary classification problems and will help to understand which attribute could effect on the result of students in the final exam. If this association is verified, at the end of this study, Measures can be taken by educational authorities in order for e-learning to achieve its goals.

3.2. Data Understanding

Exploring data is important at Data understanding stage, therefore it allows to specify potential problems in its quality [26], [27]. The conducted survey in 2021 resulting the dataset that used in this paper. It contains 2022 instances and 14 attributes, as shown in Table 1. Demographic characteristics covered such as gender .

The dataset utilized in this study came from a research survey conducted in 2021. 14 attributes and 2022 instances represent the dataset, Table 1 presents dataset attributes. The attributes cover demographic characteristics, gender, undergraduate system (Institute or College), stage, student's location, device type that used by students, Internet connection type, learning management system, use of synchronous learning, use of asynchronous learning, use of non-electronic resources, student confidence in

e-learning outcomes, The student's conviction in blended education, student contentment with e-learning, result of the final exam. Figure 2. Presents distribution of data. Google Form used to collect the data; all fields are mandatory so there is no missing data.

Table 1. Attributes

Attributes	Type of data
Gender	Text
Institute or college	Text
Stage	Text
Student's location	Text
Device type	Text
Internet connection type	Text
Learning management system	Text
Use of synchronous learning	Text
Use of asynchronous learning	Text
Use of non-electronic resources	Text
Student confidence in e-learning outcomes	Text
Student's conviction in blended education	Text
Student contentment with e-learning	Text
Result	Text

Figure 2 describe the frequency of responses for each attribute.

3.3. Data Preparation

This stage covers all of the raw data processes to build the final dataset. Inclusion and exclusion of data selection tasks included, adding new attribute possibility or amending an existing one, as well as data cleaning [14].

Machine Learning algorithms cannot be used directly on any textual data as they can only process numerical data in the form of an array. In this study, all data was in text type. Therefore, the data has been converted into numeric format to be easily handled [28].

All input attribute text values were transformed to nominal values. Each text value is simply utilized as the new attribute's nominal value. If the text attribute's value is absent, the replacement value will be as well. The dataset does not contain missing value. The Transformation process is carried out according to the rules for each attribute. Table 2 shows the transformation rules for each attribute.

If the dataset has similar number of records in both classes, equivalent importance will be given to both classes. The label attribute (result) was imbalanced, as it can be seen in Figure 2. Therefore, it was essential to utilize an oversampling method. Oversampling is technique used in data mining and data analytics to adjust uneven data classes to create balanced data sets. Gonçalves et al.[29] stated that over sampling strategies may be used to reproduce these outcomes for a more equal proportion of positive results in training . Oversampling is recommended because the amount of dataset collected is a few. Equal distribution may be accomplished by recreating the cases of the minority class using the Python Script widget.

Figure 2. Data Distribution and Representation

Table 2. Transformation Rules

Attributes	Transformation rules	Attributes type after Transformation
Gender	Male =1, Female =2	Nominal
Institute / college	Institute =1,College=2	Nominal
Stage	First=1, Second=2, Third=3, Forth=4	Nominal
Student's location	Governorate Center=1, Town=2, Countryside=3,Village=4	Nominal
Device type	Mobile=1,Tablet=2,Laptop computer=3, More than one device=4	Nominal
Internet connection type	3G=1,Wi-Fi=2,4G=3, Fiber Optic Cable=4	Nominal
Learning management system	Google Classroom=1, Moodle=2, Other=3	Nominal
Use of synchronous learning	Yes=1,No=2, Sometimes=3	Nominal
Use of asynchronous learning	Yes=1, No=2 Sometimes=3	Nominal
Use of non-electronic resources	Yes=1, No=2	Nominal
Student confidence in e-learning outcomes	Yes=1, No=2	Nominal
Student's conviction in blended education	Yes=1, No=2, Probably=3	Nominal
Student contentment with e-learning	Yes=1, No=2, To some extent=3	Nominal
Result	Pass=1 , Fail=2	Label

3.4. Modeling

In this phase, to build different DM models, the machine learning algorithms will be used. Binary classification is applied to a practical situation. The comparison operator ROC (Receiver Operating Characteristic) was used to determine the most appropriate classifiers for the problem at hand. The ROC operator allows initial filtering of the available algorithms. A ROC curve plots two parameters, True Positives Rate and False Positives Rate, at different categorization thresholds [30].

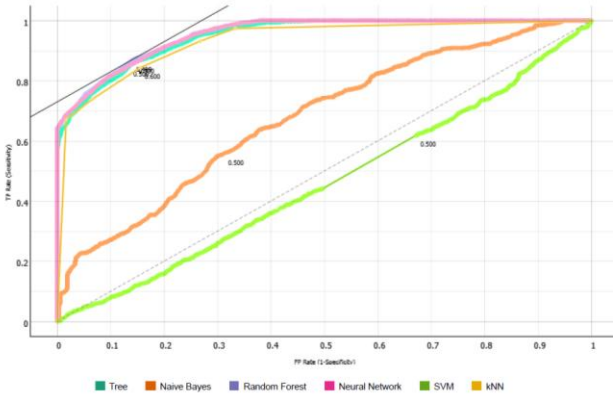

Figure 3. ROC Comparison Result

As illustrated in figure 3, ROCs attained the following best algorithms: Random Forest (RF), K-Nearest Neighbor (KNN) and Decision Tree (DT). Neural Network (NN). For each of these DM techniques, two sampling methods were tested: Split Validation with 70% of the data for training and 30% for testing and Cross Validation with 10 folds.

The weights of the attributes were assessed with the operator Weight by Correlation, It denotes the weight of each characteristic's link with the label attribute. In order to determine which qualities were more connected to the label attribute prediction, four schemas were created based on these findings. The first schema (SC1) comprises all the attributes before preprocessing while the second schema (SC2) includes all the attributes after preprocessing and the discretize continues variables filter was applied.

On the other hand, in the third scheme (SC3) only the device type, learning management system and institute or college attributes have been removed, since it had a very high correlation weight for the rating attribute Chayanukro et al.[16], it would have a greater impact on the label's predictions, which could be misleading or disguise the other features . Included attributes in SC3 are: gender, stage, student's location, internet connection type, use of synchronous

learning, use of asynchronous learning, use of non-electronic resources, student confidence in

e-learning outcomes, student's conviction in blended education, student contentment with e-learning . In the fourth schema (SC4), the device type, learning management system and institute or college attributes have been removed, discretize continues variables and oversampling method filter was applied. In total, 32 models were tested.

3.5. Evaluation

In the final stage, it's critical to evaluate the outcomes and go through the stages in depth [31]. The confusion matrix was used to evaluate the performance of all of the models that were tested Xu et al.[32], which includes the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). With these values, it is potential to obtain the metrics of accuracy (ACC), precision (P) and recall (R), and F1-score (F1) [33]. The P is the capability of the classifier not to label a positive sample that is negative. The R is the ability of the classifier to locate all the positive samples. ACC establishes the model's ability to capture true positives as being positive and true negatives as being negatives. The F1 combines the precision and recall of a classifier into a single metric by taking their harmonic mean. It is primarily used to evaluate the performance of two classifiers. We have conducted a comparison of different data mining technique for each schema, the following table show results of comparison.

According to Table 3, we see that the accuracy rate of the DM technique is between 65% and 71% using cross validation, on the other hand the accuracy rate using spit data between 50% and 65%. For SC1, the best accuracy rate was using cross validation for RF technique.

Table 3. Results of **SC1** of Each DM Technique.

Sampling Method	DM Technique	ACC	P	R	F1
Cross Validation	RF	71	71	71	71
	KNN	65	65	65	65
	DT	70	70	70	70
	NN	70	70	70	70
Split Data	RF	65	65	65	65
	KNN	64	64	64	64
	DT	50	50	50	50
	NN	63	63	63	63

Table 4 shows the accuracy rate of the DM technique are between 69% and 76% using cross validation, on the other hand the accuracy rate using spit data between 69% and 76%. For SC2, the best accuracy rate was using cross validation for RF technique, and using split data for RF and NN.

Table 4. Results of **SC2** for Each DM Technique.

Sampling Method	DM Technique	ACC	P	R	F1
Cross Validation	RF	76	76	76	76
	KNN	69	69	69	69
	DT	75	75	75	75
	NN	75	75	75	75
Split Data	RF	76	76	76	76
	KNN	69	69	69	69
	DT	75	75	75	75
	NN	76	76	76	76

Table 5 shows the accuracy rate of the DM technique are between 60% and 81% using cross validation, on the other hand the accuracy rate using spit data between 60% and 81%. For SC3, the best accuracy rate was using split data/cross validation for NN technique.

Table 5. Results of SC3 of Each DM Technique.

Sampling Method	DM Technique	ACC	P	R	F1
Cross Validation	RF	78	78	78	78
	KNN	70	70	70	70
	DT	60	60	60	60
	NN	81	81	81	81
Split Data	RF	78	78	78	78
	KNN	70	70	70	70
	DT	60	60	60	60
	NN	81	81	81	81

Table 6 shows the accuracy rate of the DM technique are between 84% and 86% using cross validation/ spit data. SC4 achieve the best accuracy rate using split data/cross validation for RF technique.

Table 6. Results of SC4 for each DM Technique.

Sampling Method	DM Technique	ACC	P	R	F1
Cross Validation	RF	86	86	86	86
	KNN	84	84	84	84
	DT	85	85	85	85
	NN	85	85	85	85
Split Data	RF	86	86	86	86
	KNN	84	84	84	84
	DT	85	85	85	85
	NN	85	85	85	85

Table 7. The Best Results of Each DM Technique.

DM Technique	Sampling Method	ACC	P	R	F1
RF	Cross Validation / Split Data	86	86	86	86
NN	Cross Validation / Split Data	85	85	85	85
DT	Cross Validation / Split Data	85	85	85	85
KNN	Cross Validation / Split Data	84	84	84	84

When looking at the results in Tables 3, 4, 5, and 6, it is clear that SC4 produced the greatest results in terms of accuracy. Because the goal property (outcome) was not balanced, this is expected behavior. With the application of over-sampling techniques, a balanced data set is obtained but with little variance in the data because the instances are replicated.

In Table 3, it is also observed that the majority of the best accuracy results used Cross Validation. This is because in the Cross Validation technique all data are utilized for training, while in the Split Validation technique only a percentage of the data is used.

On the other hand, looking at Tables 3, 4, 5, and 6, where P, R and F1 values are presented, it is noted that the results were all equivalent. This shows that the implementation of the technique was stable. The technique that achieved the best results was the RF followed by the NN. Of all the techniques, the worst was KNN, but its results were acceptable, just marginally inferior than the others.

The best models are listed in Table 7 according to their accuracy. This metric was deemed to be a better way to evaluate the models since, in

addition to predicting student achievement, we also want to find plausible factors for E-learning implementation success, that is, it is not only essential to evaluate the calculation of true positives or true negatives, but both, so that in this way it is possible to understand what is associated with the E-learning implementation success and what is not.

The attributes that contribute to building a predictive model for students participating in e-learning are: gender, stage, student's location, internet connection type, use of synchronous learning, use of asynchronous learning, use of non-electronic resources, student confidence in e-learning outcomes, student's conviction in blended education, student contentment with e-learning.

The correlations widget used as a tool for evaluating variables according to their correlation with discrete or numeric target variable, based on applicable internal scorers. In our approach, we have applied Correlation attribute evaluation, in which features are weighted and ranked based on Pearson's product moment correlation. This technique has been previously described.

The primary principle behind this method is that the relevance of a relevant feature set in a dataset may be established by assessing its correlation with the dependent variable as well as the correlation among the features. A feature set is useful for a machine learning model if the characteristics are substantially associated with the dependent class but not with each other [34].

4. Conclusion

Education sector has been severely disturbed during Covid-19 Pandemic. The WHO (World Health Organization) banned face-to-face classes. Social gatherings were also prohibited to prevent the spread of the virus. Accordingly, higher education institutes have adopted various E-learning mechanisms.

E-learning has various advantages. However, recent research stated that measuring learning outcomes and knowledge acquisition is an ongoing challenge in e-learning environment. To this end, this paper investigates different data mining techniques to predict students' performance.

CRISP-DM is used to understand and prepare the collected data. Four correlation-based attributes selection schemas are implemented to measure the significant of the selected attributes. Orange DM suite is used to construct four prediction models: RF, KNN, DT and NN. Accuracy, Precision, Recall and F1-score are calculated to measure the performance of these model. Eventually, an 86% accuracy is obtained.

Conflict of Interest

The authors declare that the publication of this article does not cause any conflict of interest.

5. References

1. T. Favale, F. Soro, M. Trevisan, I. Drago, and M. Mellia, "Campus traffic and e-Learning during COVID-19 pandemic," *Comput. networks*, vol. 176, p. 107290, 2020.
2. S.-H. Kim and S. Park, "Influence of learning flow and distance e-learning satisfaction on learning outcomes and the moderated mediation effect of social-evaluative anxiety in nursing college students during the COVID-19 pandemic: A cross-sectional study," *Nurse Educ. Pract.*, vol. 56, p. 103197, 2021.
3. A. M. Maatuk, E. K. Elberkawi, S.

Aljawarneh, H. Rashaideh, and H. Alharbi, "The COVID-19 pandemic and E-learning: challenges and opportunities from the perspective of students and instructors," *J. Comput. High. Educ.*, vol. 34, no. 1, pp. 21–38, 2022.

4. L. Yekefallah, P. Namdar, R. Panahi, and L. Dehghankar, "Factors related to students' satisfaction with holding e-learning during the Covid-19 pandemic based on the dimensions of e-learning," *Heliyon*, vol. 7, no. 7, p. e07628, 2021.
5. N. D. Oye, M. Salleh, and N. A. Iahad, "E-learning methodologies and tools," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 2, 2012.
6. K. Mukhtar, K. Javed, M. Arooj, and A. Sethi, "Advantages, Limitations and Recommendations for online learning during COVID-19 pandemic era," *Pakistan J. Med. Sci.*, vol. 36, no. COVID19-S4, p. S27, 2020.
7. S. Zarei and S. Mohammadi, "Challenges of higher education related to e-learning in developing countries during COVID-19 spread: a review of the perspectives of students, instructors, policymakers, and ICT experts," *Environ. Sci. Pollut. Res.*, pp. 1–7, 2021.
8. A. H. Al-Maqbali and R. M. Raja Hussain, "The impact of online assessment challenges on assessment principles during COVID-19 in Oman," *J. Univ. Teach. Learn. Pract.*, vol. 19, no. 2, pp. 73–92, 2022.
9. Ü. Avcı and E. Ergün, "Online students' LMS activities and their effect on engagement, information literacy and academic performance," *Interact. Learn. Environ.*, vol. 30, no. 1, pp. 71–84, 2022.
10. Z. Bilici and D. Özdemir, "Data Mining Studies in Education: Literature Review For The Years 2014-2020," *Bayburt Eğitim Fakültesi Derg.*, vol. 17, no. 33, pp. 342–376, 2022.
11. W. Xiao, P. Ji, and J. Hu, "A survey on educational data mining methods used for predicting students' performance," *Eng. Reports*, 2021.
12. J. Demšar, B. Zupan, G. Leban, and T. Curk, "Orange: From experimental machine learning to interactive data mining," in *European conference on principles of data mining and knowledge discovery*, 2004, pp. 537–539.
13. D. F. Murad, R. Hassan, W. Wahi, and B. D. Wijanarko, "A User-Item Collaborative Filtering System to Predict Online Learning Outcome," 2020.
14. T. T. Mai, M. Bezbradica, and M. Crane, "Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data," *Futur. Gener. Comput. Syst.*, vol. 127, pp. 42–55, 2022.
15. H. Zeineddine, U. Braendle, and A. Farah, "Enhancing prediction of student success: Automated machine learning approach," *Comput. Electr. Eng.*, vol. 89, p. 106903, 2021.
16. S. Chayanukro, M. Mahmuddin, and H. Husni, "Understanding and assembling user behaviours using features of Moodle data for eLearning usage from performance of course-student weblog," in *Journal of Physics: Conference Series*, 2021, vol. 1869, no. 1, p. 12087.
17. Y. Luo, N. Chen, and X. Han, "Students' Online Behavior Patterns Impact on Final Grades Prediction in Blended Courses," in *2020 Ninth International Conference of Educational Innovation through Technology (EITT)*, 2020, pp. 154–158.
18. C. Beaulac and J. S. Rosenthal, "Predicting university students' academic success and

major using random forests," *Res. High. Educ.*, vol. 60, no. 7, pp. 1048–1064, 2019.
19. P. Dabhade, R. Agarwal, K. P. Alameen, A. T. Fathima, R. Sridharan, and G. Gopakumar, "Educational data mining for predicting students' academic performance using machine learning algorithms," *Mater. Today Proc.*, vol. 47, pp. 5260–5267, 2021.
20. A. Khan, S. K. Ghosh, D. Ghosh, and S. Chattopadhyay, "Random wheel: An algorithm for early classification of student performance with confidence," *Eng. Appl. Artif. Intell.*, vol. 102, p. 104270, 2021.
21. M. Hussain, S. Hussain, W. Zhang, W. Zhu, P. Theodorou, and S. M. R. Abidi, "Mining moodle data to detect the inactive and low-performance students during the moodle course," in *Proceedings of the 2nd International Conference on Big Data Research*, 2018, pp. 133–140.
22. C. Neto, M. Brito, V. Lopes, H. Peixoto, A. Abelha, and J. Machado, "Application of data mining for the prediction of mortality and occurrence of complications for gastric cancer patients," *Entropy*, vol. 21, no. 12, p. 1163, 2019.
23. U. Thange, V. K. Shukla, R. Punhani, and W. Grobbelaar, "Analyzing COVID-19 Dataset through Data Mining Tool 'Orange,'" in *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 2021, pp. 198–203.
24. A. Hotho, A. Nürnberger, and G. Paaß, "A brief survey of text mining.," in *Ldv Forum*, 2005, vol. 20, no. 1, pp. 19–62.
25. G. Taranto-Vera, P. Galindo-Villardón, J. Merchán-Sánchez-Jara, J. Salazar-Pozo, A. Moreno-Salazar, and V. Salazar-Villalva, "Algorithms and software for data mining and machine learning: a critical comparative view from a systematic review of the literature," *J. Supercomput.*, vol. 77, no. 10, pp. 11481–11513, 2021.
26. G. D. Boca, "Factors influencing students' behavior and attitude towards online education during COVID-19," *Sustainability*, vol. 13, no. 13, p. 7469, 2021.
27. R. R. Estacio and R. C. Raga Jr, "Analyzing students online learning behavior in blended courses using Moodle," *Asian Assoc. Open Univ. J.*, 2017.
28. M. Kaushik, R. Sharma, S. A. Peious, M. Shahin, S. Ben Yahia, and D. Draheim, "A systematic assessment of numerical association rule mining methods," *SN Comput. Sci.*, vol. 2, no. 5, pp. 1–13, 2021.
29. C. Gonçalves, D. Ferreira, C. Neto, A. Abelha, and J. Machado, "Prediction of mental illness associated with unemployment using data mining," *Procedia Comput. Sci.*, vol. 177, pp. 556–561, 2020.
30. V. Giglioni, E. García-Macías, I. Venanzi, L. Ierimonti, and F. Ubertini, "The use of receiver operating characteristic curves and precision-versus-recall curves as performance metrics in unsupervised structural damage classification under changing environment," *Eng. Struct.*, vol. 246, p. 113029, 2021.
31. M. W. Rodrigues, S. Isotani, and L. E. Zarate, "Educational Data Mining: A review of evaluation process in the e-learning," *Telemat. Informatics*, vol. 35, no. 6, pp. 1701–1717, 2018.
32. J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci. (Ny).*, vol. 507, pp. 772–794, 2020.
33. Y. Rodriguez-Ortega, D. M. Ballesteros, and D. Renza, "Copy-move forgery detection (CMFD) using deep learning for

image and video forensics," *J. Imaging*, vol. 7, no. 3, p. 59, 2021.

34. S. Kumar and I. Chong, "Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states," *Int. J. Environ. Res. Public Health*, vol. 15, no. 12, p. 2907, 2018.