Modelling and Predicting Student Flexibility in Online Learning Using Machine Learning: Students' Academic Performance

Mohammed A. Aleid[1], Theyazn H.H Aldhyani [2], Osamah Ibrahim Khalaf [3]and Sameer Algburi[4]
1College of Education, King Faisal University, Al-Ahsa, 31982, Saudi Arabia, Maleid@kfu.edu.sa
2Applied College in Abqaiq, King Faisal University, Al-Ahsa 31982, Saudi Arabia, taldhyani@kfu.edu.sa
3Department of Solar ,Al-Nahrain Research Center for Renewable Energy, Al-Nahrain University, Jadriya, Baghdad, Iraq usama81818@nahrainuniv.edu.iq
4Al-Kitab University College of Engineering Techniques sameer.algburi@uoalkitab.edu.iq
Corresponding author : taldhyani@kfu.edu.sa

Abstract

With flexible learning, students are actively engaged in their own education and are held to high standards of performance. Online academic courses make it easier for students to receive personalized education because they provide students with more flexibility to concentrate on what is most important to them and give them greater control over their own education. This study's objective was to investigate whether there is a correlation between how well students succeed in online classes and the extent to which they make use of the schedule and the geographical and resource flexibility offered by such programmes. This article uses a developing approach for predicting and classifying the flexibility in online learning of students who are at risk of failing due to academic and demographic variables. The K-nearest neighbors (KNN) method, the random forest (RF) method, and the logistic regression method were used to categorise the students participating in flexible online learning. The information for the dataset came from Kaggle, and it was gathered for use in testing machine learning. The dataset had a total of 1,875 instances representing 11 different features. Also, accuracy, precision, sensitivity and f-score metrics were applied to evaluate the system. The results show that the RF algorithm has a high accuracy percentage of 85%. The empirical findings demonstrate that students formed distinct patterns of learning time, location and access to knowledge. This suggests that flexibility was used to a significant degree. Patterns in learning time and the availability of learning materials were shown to have a substantial relationship with the accomplishments of the students. Gaining an understanding of the various patterns of flexibility utilisation has the potential to promote the development of tailored learning and to improve cooperation among students who share comparable traits.

Keywords: flexible learning, online academic course, machine learning, online learning

1. Introduction

The ways in which we interact with one another and learn and share information may be profoundly affected by digital technology. It can seem that using technology in the classroom is a new development in our culture. However, such use in mobile and remote settings is nothing new. Distance education, online learning, human–computer interaction and computer-supported collaborative learning have all contributed to the fast-expanding corpus of research that aims to understand the experiences and behaviours of students and teachers in technology-assisted learning settings [1]. The current state of knowledge shows that using digital technology successfully in remote and adaptable settings is a difficult task. The tradition of flexible and remote education goes back many years. A large amount of research has produced insights about learning that differ from those attained in more conventional face-to-face settings, such as classrooms or lecture halls. In turn, this expanded understanding has led to the development of novel pedagogies and ideas that aid educators in meeting the requirements of students in nontraditional classroom settings. In what follows, we will briefly examine the evolution of distance and flexible learning, some of the most important frameworks in the field, some of the most influential theories and pedagogies, the technological affordances that support distance and flexible learning and some of the challenges that come along with these approaches. Among the many potential advantages of distance learning is that it removes the constraints of time and place from the classroom (e.g., classroom or campus)[2,3]. This means that students may study wherever and whenever they choose (at home, at the workplace, at school, at university, in a café, etc.) and at their own speed without having to physically meet with an instructor. Increased accessibility is another perk of online courses. People who were previously unable to engage in formal education because of factors such as their location, personal circumstances, financial restraints, disability or the lack of available courses now have greater access to these options [4]. Today, when people think of 'distance learning', they usually picture a course that takes place entirely online.

Students have more options available to them in terms of time, location and the rate at which they study when they participate in e-learning. Students have the potential to participate in the e-learning process in a meaningful way thanks to the technologies that support e-learning [5].However, for students to achieve their educational objectives via e-learning and other forms of remote learning, they must take responsibility for the pace and structure of their online learning experiences and act with greater independence [6]. This raises problems regarding how to create an effective e-learning environment for students. Individual variations in student learning have a major influence on learning outcomes, which in turn leads to shifts in pedagogical methods according to the findings of research published in the academic literature on the efficacy of e-learning [7]. Therefore, to create successful online learning experiences, it is essential to consider the unique characteristics of each student and to use individualized instructional strategies [8]. Within the confines of this discussion, adaptability and adaptable learning stand out as phenomena of a more general nature. Flexible learning allows for further individualisation in terms of what, when and how one learns. Flexibility in the classroom is based on the idea that students should have input into how they learn. The time and location of lessons, the materials and methods used to teach them, the prerequisites for participation, the available technological tools and the means of communication are just a few examples of where adaptability may be found. It caters to students' individual interests and demands by providing a variety of study options. Flexible learning relies heavily on the use of technology, yet it goes well beyond the use of technology in education. It is also important to consider the pedagogy behind flexible learning, the tactics employed to bring additional freedom for learners and the institutional structures utilised to enable the provision of flexible learning alternatives within an organisation.Flexible learning and remote learning are assumed to be intertwined for the purposes of this article, with distance learning serving to enable and improve flexible learning. Online learning flexibility is presented in Figure 1.

Figure 1. Online learning flexibility

With the convenience of online classes students' individual needs are considered in flexible learning by allowing them to provide input in how, when, where and why they acquire knowledge [9]. As a result, the requirements, interests, backgrounds and learning styles of individual students are prioritized in a flexible learning environment. This is an indication that we are moving away from outdated teaching methods in favour of fresh, student-focused ones [10]. Research has shown that students' needs for flexibility are a driving factor in their decision to enrol in online courses [11]. However, just because learners are given a variety of possibilities for flexible learning does not mean that they will inevitably engage in deep learning [12]. There is a great deal of responsibility that comes with this liberty [13]. Therefore, in a flexible learning environment, students must take more initiative, exercise more autonomy and devote more time and energy to their studies [14]. Research suggests that pupils need constant guidance during this process [15].

To discover how to effectively use flexibility and support students' learning, it is crucial to examine how students utilise it and how it connects to their accomplishments. A growing field of study, educational data mining (EDM) follows students' digital footprints to better understand how they learn and interact with the world. Learning management systems (LMSs) and other forms of digital education keep track of student progress in a variety of ways [17-20].Thus, decisions can be informed by data rather than relying entirely on students' perceptions of their own performance, and instructional tactics can be fine-tuned accordingly). Only a small number of studies have utilised LMS data to examine how students make use of online courses' adaptive features and how this relates to their final grades. Using interviews and an analysis of course data, reported three approaches to encouraging students' use of adaptable online materials.

In recent years, many scholars have been interested in learning analytics (LA), also known as EDM [3], which is defined as the process of analysing and discovering patterns in learners' data for decision-making reasons. With the use of learning analytic technologies, institutions may learn more about their students' current positions, behaviours and preferences in relation to their peers and the desired learning outcomes. Because of this, content can be adapted for each student depending on their own goals and preferred methods of learning. Data from a wide variety of sources, such as student enrolment information, past and present academic records, student online behaviour, student surveys via questionnaires concerning courses and teaching techniques and data from online discussion forums, are used by LA systems to evaluate students' learning behaviours in online education. Researchers have examined several aspects of students' learning behaviours to see how they might best anticipate and improve students' performance and retention in class. Researchers have investigated many different machine-learning models, such as support vector machines (SVMs), linear regression (LR), RF and deep learning models, such as convolutional neural networks (CNNs) and long short-term memory (LSTM), to predict and analyse students' outcomes in online courses [21-25]. This research makes a significant contribution to our knowledge of the role that flexibility dimensions play in online learning experiences and the results of those experiences. The practical ramifications of these studies' findings, as well as ideas for future researchers, are discussed in the study background section.

2. Contributions

In the context of remote learning, the concept of flexibility requires additional investigation from a range of perspectives. Although numerous studies have shown the significance of adaptability in e-learning, the direct effect that students' perceptions of adaptability have on their behavioural engagement and academic success is not yet fully understood. Also, earlier research indicated that more studies were required to investigate the possibility of a causal connection between the various elements of flexible learning and the results of learning when applied to the setting of open and remote learning. As a result, the goal of this research is to investigate how a perceived environment's level of flexibility affects a learner's level of behavioural engagement and academic achievement in an online setting. To the best of knowledge, no prior research has studied the impact of perceived flexibility characteristics on the behavioural engagement of learners, as determined by objective measurements and academic achievement. It is intended that this study will make a significant addition to research within the area of remote learning and will expand our knowledge of the function that flexibility dimensions play in the learning of students in an e-learning setting. Machine learning strategies were used to categorise the adaptability of online education. Also, an innovative approach was used in the present research to quantify the behavioural involvement of students using data from LA indicators. One of the novel parts of the research is how the authors used the students' experiences with online learning to determine their level of behavioural engagement.

3. Background of the Study

Today, both traditional classrooms and computer-based courses—commonly referred to as 'e-learning' [26]—are common in the academic world. The world's schools are quickly adopting e-learning because it offers the opportunity to use cutting-edge, potentially more effective teaching methods. Especially in the post-COVID-19 era, advancements in information and communication technology (ICT) have been essential in expanding web-based pedagogical practices [27]. E-learning technologies have not only been crucial in supporting in-person student–teacher conferences [28], but have also become an essential aspect of online teaching. Many obstacles have arisen because of the shift from conventional classrooms to online learning, with students' lack of engagement being especially detrimental to their academic outcomes. Therefore, it is crucial to create methods that can diagnose causes and predict future academic outcomes for children. This has been possible because of the abundance of recent research [29–35] that has investigated e-learning spaces. To classify a student as either a poor or a high achiever in the classroom, Brahim [36] generated an 86-dimensional feature space from which only useful characteristics were explored using several machine learning methods. The suggested approach was tested in three experimental settings, where the author found 97.4% accuracy using an RF classifier. To assess the efficacy of their own strategies, Nikola et al. [37] used numerous machine learning techniques. Both classification and regression tasks were used by the authors to predict the exam results. Classification included assigning students to one of two groups—'pass' or 'fail'—while regression attempted to predict each student's actual test results. Sekeroglu et al. [38] employed a variety of machine learning algorithms to examine the Student Performance Dataset and the Students Academic Performance Dataset, the former of which was used for prediction and the latter for categorization. To foretell how well students will perform utilising an e-learning system, Burgos [39] considered their online behaviour. Using information gleaned from students' Sakai platform log-in histories and the LMS, the author classified them according to their preferred learning methods [40]. Preprocessing, feature selection and parameter optimisation were carried out before classification. Classifying students in this way aids in estimating what they will do in a certain course. Additional research [41] has shown the usefulness of using a student's past grades in conjunction with machine learning methods to forecast their future performance. To assist students in avoiding making hasty judgments about leaving a course, a dashboard was developed to provide real-time predictions of their performance. Another study [42] employed machine learning to predict students' participation based on behavioural variables and then analysed the impact of engagement on evaluation scores. A dashboard that shows student actions in the learning environment might assist teachers in quickly seeing pupils who are not actively participating. To better engage students in the classroom and, by extension, improve their performance, an adaptive gamified learning system [43] was designed that combined gamification with EDM. In an electronic classroom setting, the efficiency of gamification was measured against that of adaptive gamification.

The probabilistic semantic-based indexing model (PLSI) [43] and latent dirichlet allocation (LDA) [44] are two popular methods used to find hidden themes in text. In the PLSI, a low-dimensional representation is used to assess similarities, which helps highlight the subject based on semantic indexing. Knowledge discovery and documentation have made use of the LDA model, which is an unsupervised model. Several methods, all based on the LDA framework, have been suggested to mine web material and find themes for online communities. Using deep learning techniques, Nagori et al. created a system to provide text-based recommendations for use in online education. By developing similarity metrics, they were able to use the topic model [45]. The quality of online student comments was modelled by Zhong et al. [46], who created a collection of subject variables that were highly correlated with comment quality. The data sparsity issue [47] arises when these models are applied directly to short texts. The biterm topic technique (BTM) was introduced by Yan et al. [48] as a method for detecting themes in small amounts of textual data; their model created the term co-occurrence over the whole corpus to facilitate better topic learning.

Education that takes place at a distant location is called distance education. Understanding students' learning environments may be facilitated by the analysis and mining of data produced in an online education setting. Structured data, such as student performance and activity or course discussion forums, have been analysed using data-driven methodologies [49]. To improve the quality of data mining in remote learning, many different studies have been conducted and online platforms have been used to promote the flow of information regarding online education. With the use of text mining and regression approaches, Kagklis et al. [50] were able to enhance the classroom experience and predict student outcomes. Rooyen surveyed accounting majors at the University of South Africa to get their perspective on whether they thought social networking applications would be useful in the classroom. Most students surveyed [51] had a favourable impression of utilising social media for schoolwork. In [52], the authors suggested an online forum for the discussion and sharing of research on artificial immune systems. The technology facilitates on-demand training that helps students acquire relevant competencies in a timely manner.

4. Materials and Methods

4.1 Proposed Framework

It is essential to have flexible learning that includes e-learning, open and remote learning and blended learning. With the rising prevalence of technology applications in education, the phrases e-learning and flexible learning are sometimes used interchangeably (Li & Wong, 2018). Because of the impact it has on both the learning experiences and the results of online education, flexibility is rapidly emerging as one of the most important tools in the field of e-learning. The proposed framework for classifying flexible online learning using various machine learning approaches is presented in Figure 2.

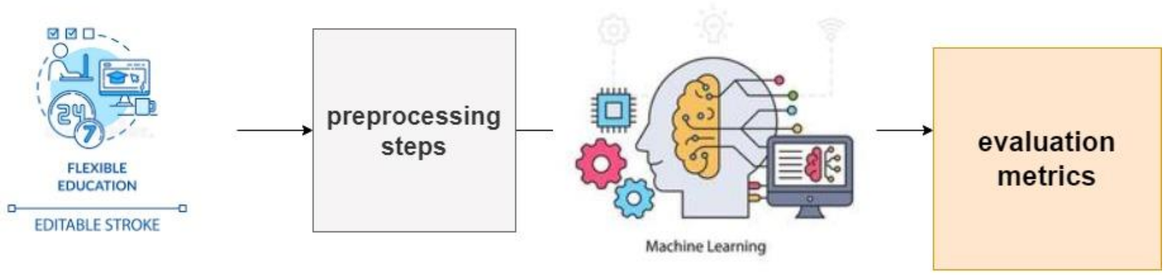

Figure 2. Framework of the proposed system

4.2 Dataset

The dataset was collected from the Student Flexibility in Online Learning dataset through the Kaggle website. The dataset contains 1,206 instances. The dataset contains the academic information for students from the following levels: students in public and private schools, colleges and universities. Figure 3 shows the features of the data.

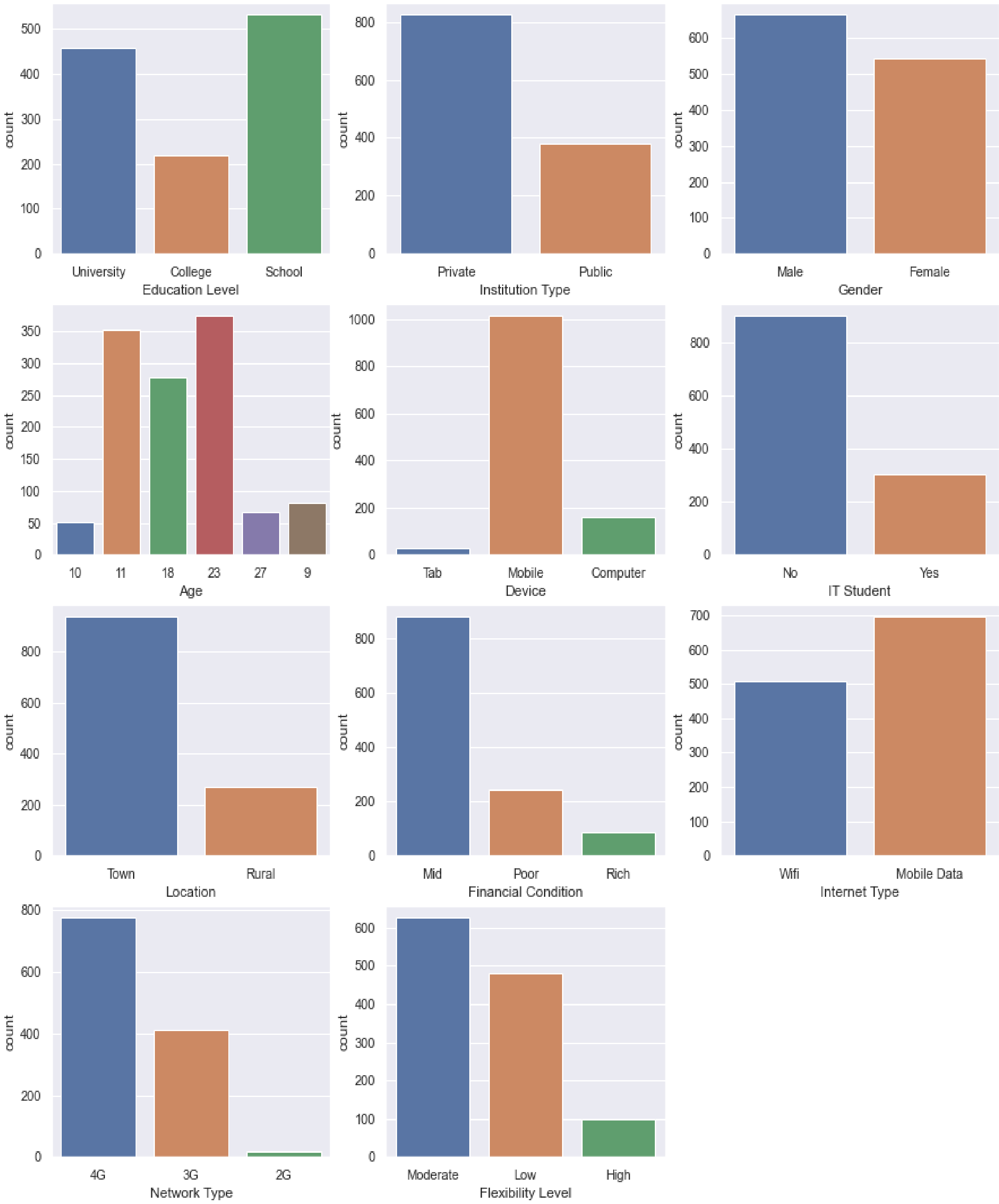

Figure 3. Data features

4.3 Data Preprocessing

Preprocessing is a key stage in data mining, so that the data may be utilised by algorithms in their final form. The dataset underwent three primary preparation operations: data cleaning, features encoding and features scaling. Python was used for the language, while Excel was used for preprocessing. Figure 4 shows the preprocessing steps for predicting the students' flexibility in online learning.

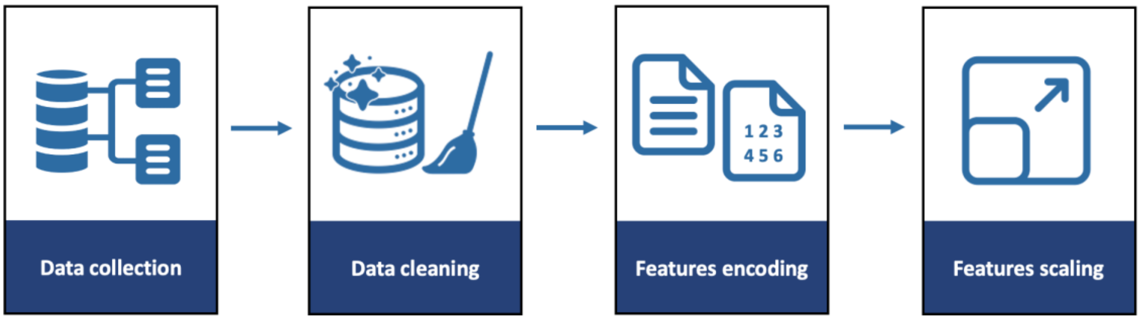

Figure 4. Preprocessing steps

The data we collect are often cluttered and disordered. Through the addition of missing values and the elimination of noise, outliers may be corrected in the data throughout the cleaning process. Several students in our dataset failed to complete the survey and this was treated as missing data. The demographics of the dataset are shown in Figure 5.

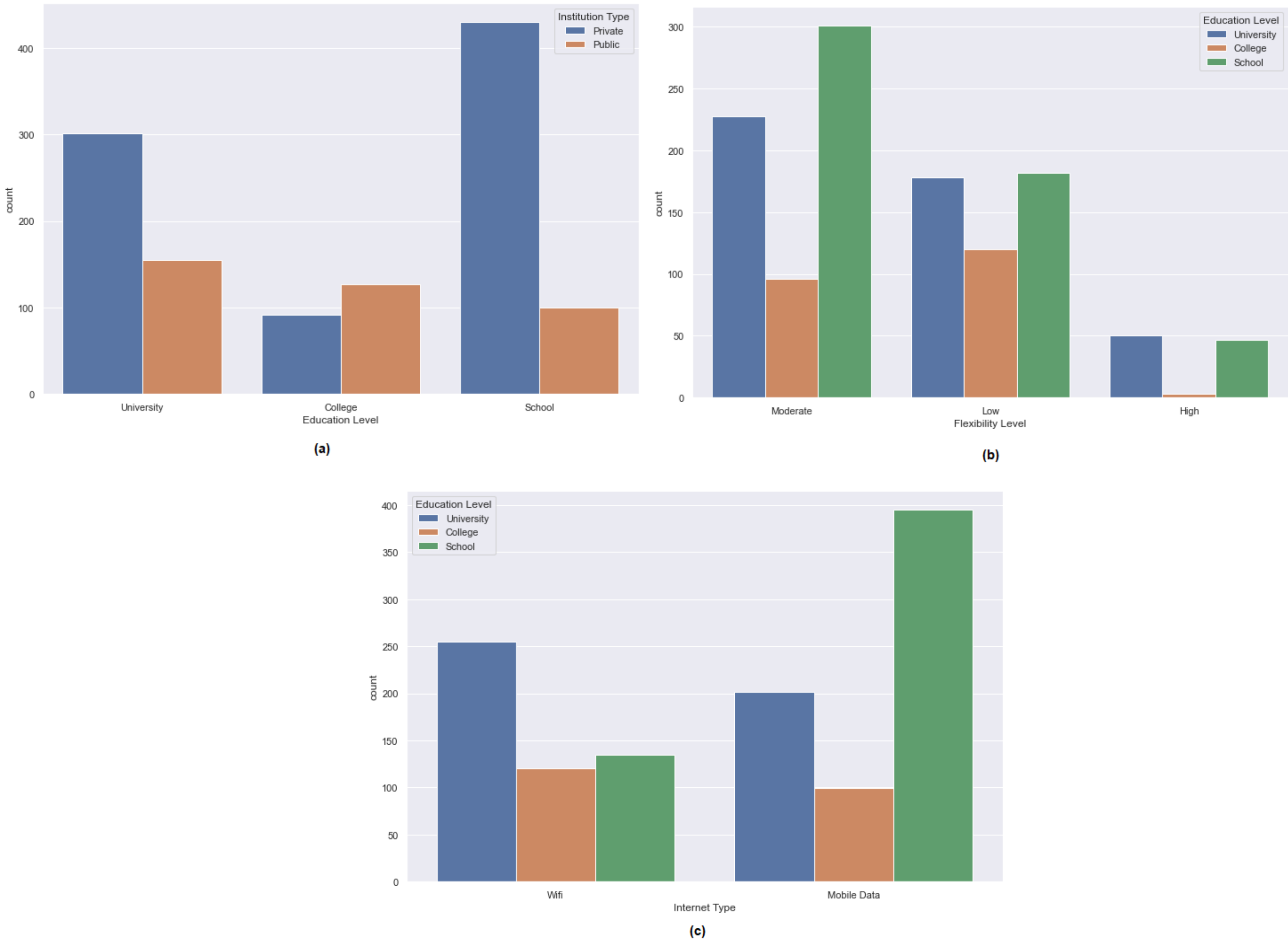

Figure 5. Demographics

4.4 Features Encoding

During data cleansing, anomalies may be fixed by filling in missing values and removing unwanted noise. The data contain all categorical variables about the students' flexibility in online learning. The encode function was applied to convert the categorical variable into a numerical variable.

4.5 Normalisation Method

By rescaling the data from a larger range, say 1.0 to 0.0, feature scaling normalizes a collection of independent variables or data features. In certain cases, this may shorten the training time for algorithms and minimise their mistake rates. The MinMaxScaler method was used for scaling features.

4.6 Correlation

The degree to which changes in the value of one variable may be used to anticipate changes in the value of another variable is measured by a statistical concept known as the correlation coefficient. In variables that have a positive correlation, the value rises or falls along with the other variables. The correlation coefficient is presented in Figure 6.

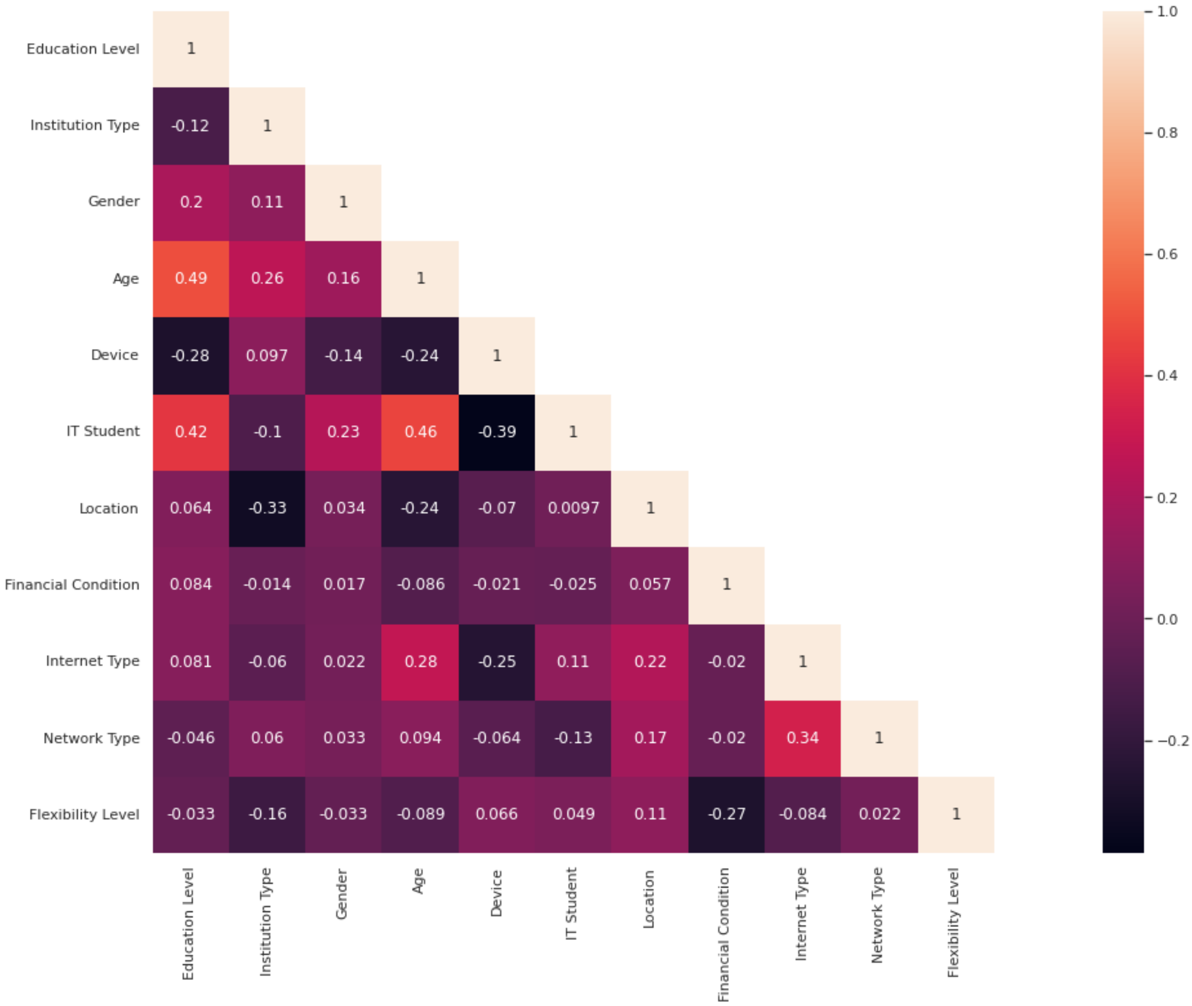

Figure 6. Correlation coefficient plot

4.7 Machine Learning Algorithms

4.7.1 RF Approach

The application of the machine learning strategy known as RF can help solve problems relating to regression and classification. An approach known as ensemble learning, which employs several classifiers to solve a problem, is utilised. The RF algorithm comprises numerous different decision trees. Bagging or bootstrap aggregation is utilised to train the 'forest' that is produced by the RF approach. The performance of machine learning ensembles can be improved with the use of a meta-algorithm called bagging. The RF method decides the outcome after considering the projections made by the decision trees. It does this by taking the findings of many trees and averaging them. The use of a greater number of trees results in an improvement in accuracy. The efficiency of the algorithm may be diminished if the same data are fed into each branch of the decision tree. When making predictions using a decision tree, if just a small sample of the dataset is used, there will be a significant amount of variation.

It is possible that different features of the dataset do not impact a model's overall prediction of RF. It has been shown that RF algorithms are more accurate than decision trees. More research has indicated that the RF method's superior prediction accuracy is due to its use of several decision tree outputs in the forecasting process. In both decision trees and random forests, information gain is the primary determinant of feature importance.

4.7.2 K-Nearest Neighbors (KNN)

The KNN algorithm is a supervised machine learning technique that can be utilised to solve predicting problems in either the regression or classification domains. However, the most important use can be found in the realm of predictive classification problems. The KNN algorithm gives a value to a new datapoint based on how closely it resembles the points in the training set. This idea is referred to as 'feature similarity', and it is used by the algorithm.

$$\mathrm{E_i} = \sqrt{(\mathrm{c_1} - \mathrm{c_2}) + (\mathrm{d_1} - \mathrm{d_2})^2} \tag{1}$$

where c_1, c_2, d_1, and d_2 are the input data variables.

4.7.3 Logistic Regression

To classify data, the statistical approach known as logistic regression should be used if the dependent variable in the machine learning model is binary. Logistic regression is a statistical technique that may be used to describe the data, as well as the relationship between a single dependent variable and numerous independent variables. There are three possible types of values for indirect variables: nominal, ordinal and interval. The statistical technique known as logistic regression got its name from the concept of a logistic function. This helpful mathematical tool also goes by the names sigmoid function or inverse logistic function. This logistic function can represent any number in the range of 0 to 1, inclusive.

$$S(x) = \frac{1}{1+e-x} \tag{2}$$

An integer is sent through the sigmoid function, which returns a 0–1 as the outcome of the operation. The sigmoid function returns the likelihood of categorisation in each case. When S(x) is less than 0.5, the data are classified as class A, and when S(x) is more than 0.5, the data are classified as class B.

4.8 Performance Metrics

In this study, sensitivity, specificity, accuracy, recall and the F1 score were used to determine the system's efficacy.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (3)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$\text{Fscore} = \frac{2 * precision * \text{sensitivity}}{precision + \text{sensitivity}} \times 100\% \quad (6)$$

where y (i,exp) is input data from the flexibility online learning, and y (i,pred) is the output of the developing system, and where the confusion metrics of the flexibility online learning system, such as true positive (TP), true negative (TN), false positive (FP) and false negative (FN), were utilised as parameters for examination of the model.

5. Experiment

In this section, we present the classification performances of the RF, KNN and logistic regression approaches for evaluation metrics such as accuracy, precision, recall and f-score. In this experiment, we carried out 10 significant ant features to determine the student flexibility of online learners. Three classes, namely moderate, low and high, were considered according to the features of students. A detailed experiment for this study is presented in the following subsection.

5.1. Experiment Setup

Scikit-learn was used to implement the RF, KNN and logistic regression. Two types of classifications were employed to examine the shows' contents. The hardware configuration for this programme included a 3.20 GHz Intel (R) Core (TM) i7-4770 processor, 8 GB of RAM and a 64-bit version of Windows 10.

5.2. Splitting Data

The data set was split between a 70% training portion and a 30% test portion. The results from both ML methods were tested. The sizes of the datasets are shown in Table 1.

Table 1. dataset

Variable	Training size	Testing size
Dataset	1256	619

6. Results

In this section, the results of machine learning to classify students' online learning flexibility are presented. Table 2 shows that the RF algorithm achieved 85% accuracy metrics. This accuracy was high compared with different existing algorithms. It is noted that the RF algorithm attained a high percentage in the moderate class.

Table 2. Results of RF algorithm to classify student's online learning flexibility.

Algorithms	Classes	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
SVM	Moderate	85	90	95	**93**
	Low		88	79	**83**
	High		78	81	**79**
	Weighted Average		**85**	**85**	**85**

Table 3 shows the results of the KNN algorithm to predicting the students' flexibility in online learning. The KNN method achieved an accuracy of 78%, whereas we observed that the KNN had a high percentage in the moderate class.

Table 3. Results of the KNN algorithm to classify students' online learning flexibility.

Algorithms	Classes	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
SVM	Moderate	78	87	90	**88**
	Low		76	79	**77**
	High		70	65	**67**
	Weighted Average		**78**	**78**	**78**

The results of regression logistic methods is presented in Table 4, it is observed that the RL methods was scored 65% with terms of accuracy metrics. The RL method achieved very low accuracy compared with the RF and KNN algorithms. Also, the RL method achieved high accuracy in the moderate class.

Table 4. Results of RL algorithm to classify students' online learning flexibility.

Algorithms	Classes	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
SVM	Moderate	65	71	69	**70**
	Low		65	61	**63**

High	60	65	**62**
Weighted Average	**65**	**65**	**65**

6.1 Results Discussion

Some students are left out of the digital learning experience because of access and pricing barriers, which may shed light on broader issues of social, economic and cultural inequality. Over time, the problem of low retention rates in online and hybrid classes has persisted (as compared to the same classes taught in person). Students' decisions to continue or not with distance learning courses are influenced by numerous factors, including feelings of isolation, technical frustrations, a lack of requisite prior experience, inadequate technical support, unwilling participation from faculty or students, a lack of timely feedback, time constraints, other responsibilities and low learner motivation.

However, standard classroom instruction cannot accommodate a wide variety of student learning styles. As a result, institutions may evaluate the results of implementing new teaching methods and adapting how they offer material to students. Making students responsible for their own education and encouraging them to try different approaches to learning helps transfer the onus of learning from the teacher to the student. Students who take greater ownership of their education tend to be more committed to it.

This research employed educational data machine learning techniques to analyse students' use of course-specific flexibility in resources and access to learning materials across four online academic courses. The students' adaptability in their online learning was measured using machine learning algorithms, including RF, KNN and LR. Based on the subjects studied, this study classified students' use of flexibility in online learning into three categories: moderate, low and high. The findings not only showed that students made extensive use of flexibility but also provided evidence of the many ways in which they did so.

In terms of online learning flexibility, most students in this survey obtained most of their education outside traditional classroom settings and spread their studying out equally throughout the seven days of the week. The vast majority of today's students are urbanites and middle-class earners who want to pursue their education over the Internet. When it comes to mobile technology, most students nowadays rely on the 4G Internet on their smartphones. Students' access to online learning is greatly influenced by factors such as location and time. The RF tree algorithm achieved a high classification accuracy of 85%. Figure 3 shows the performance of the proposed machine learning method in terms of accuracy metrics.

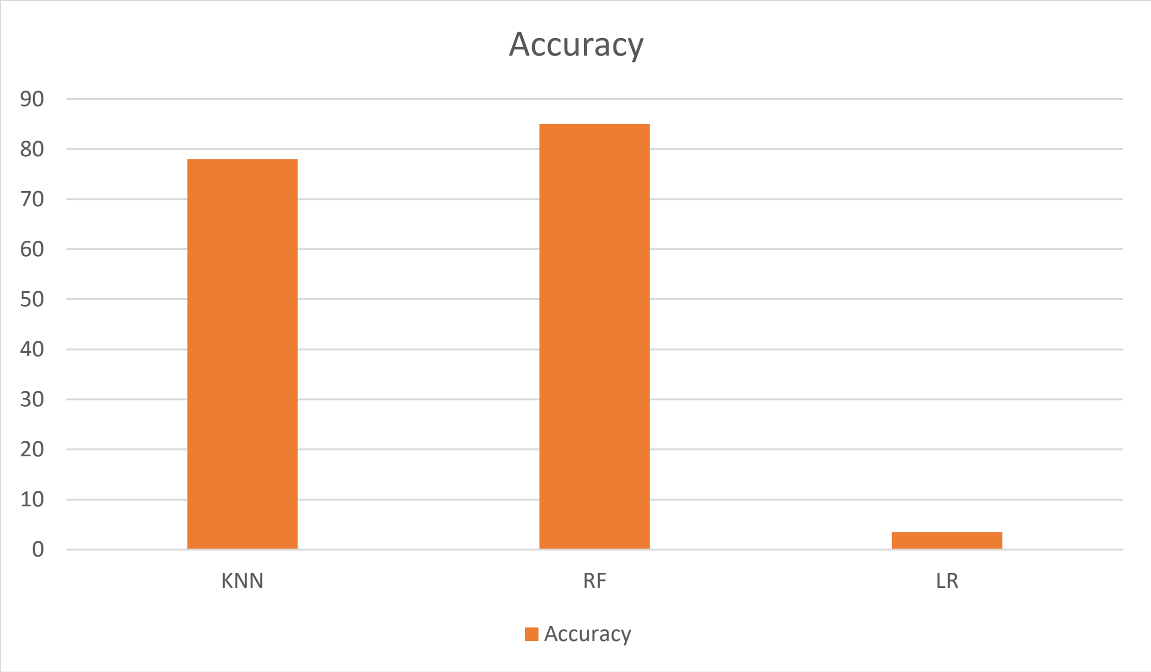

Figure 7. Performance of machine learning model

When compared to online learning environments, which provide students with greater leeway in terms of how and when they access learning materials, conventional classroom settings, in which students are expected to follow a predetermined curriculum, offer less room for manoeuvre [53]. The findings from an analysis of how often students accessed the learning materials showed that they often made use of this option.

Students can develop greater self-discipline and life skills while taking advantage of the flexibility of online learning, both of which will be beneficial to them when they enter the workforce in the future. Because of the need to regularly attend courses, students in conventional education systems do not sufficiently prioritise the process of cultivating their own personal development. Students who can show that they have obtained the aforementioned abilities via the use of online learning may eventually find themselves in a position inside an organization that is more senior than their current one. Therefore, flexibility in and of itself has a significant influence on the way pupils acquire knowledge, and its allure is not without foundation. It is essential to an individual's achievement to provide them with the freedom to choose the educational path that best suits them.

7. Conclusion

Most students who choose an online degree programme over a traditional one do so because of the freedom it provides them. Despite its apparent drawbacks, flexibility has been shown to have a favourable effect on students' academic performance. Institutions are starting to include online learning components after swiftly realizing their usefulness and influence. Not only does this keep students actively involved, it also frees up valuable class time by relieving instructors of the burden of assigning work. In addition to classroom participation, some courses require students to take part in online quizzes and discussions, which may encourage more interaction and a deeper understanding of the course material.

In conclusion, this study's findings indicate that most students use many strategies for making the most of learning flexibility in terms of time, place and access to learning resources. As a result, it is crucial to provide students with some degree of autonomy in designing their own online learning experiences to best suit their needs. Also, this study's results linked students' adaptability with higher GPAs. Both the amount of time and effort a student puts into their studies throughout the course of the semester and their level of interest in the topics being taught in class have a direct bearing on their final grades. This study contributes to the literature on adaptable learning by using EDM to assess student behaviour and to provide evidence of how students make use of flexibility. In the real world, it might be used to aid in the development of adaptive online courses that are tailored to the specific requirements of individual students. Achieving this aim may require teachers to offer scaffolding for students to make more strategic use of learning resources (such as assignments) throughout the semester, especially in the last few weeks of class. Also, this study's findings may make it easier for students who have similar characteristics to get to know one another and work together (e.g., the same learning time, course content).

Funding: This research and the APC were funded by the Deanship of Scientific Research at King Faisal University for the financial support under grant No. 702

Conflict of Interest: The authors declare that they have no conflict of interest

References

1. Baker, R. & Siemens, G. (2014) Educational data mining and learning analytics. In K. Sawyer (Ed.), Cambridge h andbook of the l
2. earning s ciences (2nd ed., pp. 253–274). New York, NY: Cambridge University Press.
3. Bates, T.(2001). National strategies for e-learning in post-secondary education and training, In Fundamentals of educational planning, 70 (pp. 1–134). Paris: UNESCO
4. Moore, M. and Kearsley, G. (2012) Distance Education: A Systems View of Online Learning. 3rd Edition, Wadsworth, Belmont.
5. Kim, H.J., Hong, A.J. & Song, HD. The roles of academic engagement and digital readiness in students' achievements in university e-learning environments. Int J Educ Technol High Educ 16, 21 (2019).
6. Shearer, R. L., & Park, E. (2018). Theory to practice in instructional design. In M. G. Moore & W. C. Diehl (Eds.), Handbook of Distance Education (4th ed.), (pp. 260 -280). New York: Routledge.
7. Dabbagh, N. (2007). The Online Learner: Characteristics and Pedagogical Implications. *Contemporary Issues in Technology and Teacher Education, 7*(3), 217-226. Waynesville, NC USA: Society for Information Technology & Teacher Education. Retrieved January 8, 2023 from.
8. Collis, B., & Moonen, J. (2001). Flexible Learning in a Digital World: Experiences and Expectations. Kogan Page.

9. Bergamin, P. B., Ziska, S., Werlen, E., & Siegenthaler, E. (2012). The relationshipbetween flexible and self-regulated learning in open and distance universities.The International Review of Research in Open and Distributed Learning,13(2), 101–123. doi: https://doi.org/10.19173/irrodl.v13i2.1124
10. Boer, W. D., & Collis, B. (2005). Becoming more systematic about flexible learning: Beyond time and distance. ALT-J, 13(1), 33-48. doi.org/10.1080/0968776042000339781Brown, S., & Smith, B. (2013). Research, teaching and learning in higher education. London: Routledge. doi:https://doi.org/10.4324/9781315041568.
11. Gillingham, Margaret and Carol A. Molinari. "Online Courses: Student Preferences Survey." (2012).
12. Collis, B., & Moonen, J. (2001).Flexible learning in a digital world: Experiences and expectations. London: Routledge. doi: https://doi.org/10.4324/9780203046098
13. Collis, B., & Moonen, J. (2002). Flexible learning in a digital world.Open Learning 17(3), 217–230.doi: https://doi.org/10.1080/0268051022000048228
14. Collis, B., Vingerhoets, J., & Moonen, J. (1997). Flexibility as a key construct in European training: Experiences from the TeleScopia Project.British Journal of Educational Technology, 28(3),199–217. doi: https://doi.org/10.1111/1467-8535.00026
15. Cornelius, S., & Gordon, C. (2008). Universalists, butterflies and changelings: Learners' roles and strategies for using flexible online resources. Proceedings of EdMedia: World Conference on Educational Media and Technology, 2008 (Vol. 1), 4052–4057.
16. Cornelius, S., Gordon, C., & Ackland, A. (2011). Towards flexible learning for adult learners in professional contexts: An activity-focused course design. Interactive Learning Environments, 19(4), 381–393. doi: https://doi.org/10.1080/10494820903298258.
17. Hamim, T.; Benabbou, F.; Sael, N. Survey of Machine Learning Techniques for Student Profile Modelling. *Int. J. Emerg. Technol. Learn.* 2021, *16*, 136–151.
18. Prenkaj, B.; Velardi, P.; Stilo, G.; Distante, D.; Faralli, S. A survey of machine learning approaches for student dropout prediction in online courses. *ACM Comput. Surv. (CSUR)* 2020, *53*, 1–34.
19. Gardner, J.; Brooks, C. Student success prediction in MOOCs. *User Model. User-Adapt. Interact.* 2018, *28*, 127–203.
20. Katarya, R.; Gaba, J.; Garg, A.; Verma, V. A review on machine learning based student's academic performance prediction systems. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 254–259.
21. Kuzilek, J.; Hlosta, M.; Zdrahal, Z. Open university learning analytics dataset. *Sci. Data* 2017, *4*, 170171.
22. Adnan, M.; Habib, A.; Ashraf, J.; Mussadiq, S.; Raza, A.A.; Abid, M.; Bashir, M.; Khan, S.U. Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. *IEEE Access* 2021, *9*, 7519–7539.
23. Jha, N.; Ghergulescu, I.; Moldovan, A. OULAD MOOC Dropout and Result Prediction using Ensemble, Deep Learning and Regression Techniques. In Proceedings of the 11th International Conference on Computer Supported Education, Heraklion, Greece, 2–4 May 2019; SciTePress: SetÃžbal, Portugal, 2019; Volume 2, pp. 154–164.
24. Song, X.; Li, J.; Sun, S.; Yin, H.; Dawson, P.; Doss, R.R.M. SEPN: A Sequential Engagement Based Academic Performance Prediction Model. *IEEE Intell. Syst.* 2021, *36*, 46–53.
25. Hussain, M.; Zhu, W.; Zhang, W.; Abidi, R. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. *Comput. Intell. Neurosci.* 2018, *2018*, 6347186.
26. Manjarres, A.V.; Sandoval, L.G.M.; Suárez, M.S. Data mining techniques applied in educational environments: Literature review. *Digit. Educ. Rev.* 2018, *33*, 235–266.

27. Zareie, B.; Navimipour, N.J. The effect of electronic learning systems on the employee's commitment. *Int. J. Manag. Educ.* 2016, *14*, 167–175.
28. Muniasamy, A.; Alasiry, A. Deep Learning: The Impact on Future eLearning. *Int. J. Emerg. Technol. Learn.* 2020, *15*, 188–199.
29. Alharthi, A.D.; Spichkova, M.; Hamilton, M. Sustainability requirements for elearning systems: A systematic literature review and analysis. *Requir. Eng.* 2019, *24*, 523–543.
30. Umer, R.; Susnjak, T.; Mathrani, A.; Suriadi, S. On predicting academic performance with process mining in learning analytics. *J. Res. Innov. Teach. Learn.* 2017, *10*, 160–176.
31. Lu, O.H.; Huang, A.Y.; Huang, J.C.; Lin, A.J.; Ogata, H.; Yang, S.J. Applying learning analytics for the early prediction of Students' academic performance in blended learning. *J. Educ. Technol. Soc.* 2018, *21*, 220–232.
32. Widyahastuti, F.; Tjhin, V.U. Performance prediction in online discussion forum: State-of-the-art and comparative analysis. *Procedia Comput. Sci.* 2018, *135*, 302–314
33. Zhang, W.; Huang, X.; Wang, S.; Shu, J.; Liu, H.; Chen, H. Student performance prediction via online learning behavior analytics. In Proceedings of the 2017 International Symposium on Educational Technology (ISET), Hong Kong, China, 27–29 June 2017; pp. 153–157.
34. Koutina, M.; Kermanidis, K.L. Predicting postgraduate students' performance using machine learning techniques. In *Artificial Intelligence Applications and Innovations*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 159–168.
35. Alzahrani1, N.A.; Abdullah, M.A. Student Engagement Effectiveness in E-Learning System. *Biosci. Biotechnol. Res. Commun. Spec. Issue Commun. Inf. Technol.* 2019, *12*, 208–218.]
36. Brahim, G.B. Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features. *Arab. J. Sci. Eng.* 2022, 10225–10243
37. Tomasevic, N.; Gvozdenovic, N.; Vranes, S. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Comput. Educ.* 2020, *143*, 103676.
38. Sekeroglu, B.; Dimililer, K.; Tuncal, K. Student performance prediction and classification using machine learning algorithms. In Proceedings of the 2019 8th International Conference on Educational and Information Technology, Cambridge, UK, 2–4 March 2019; pp. 7–11.
39. Burgos, D. Background similarities as a way to predict students' Behaviour. *Sustainability* 2019, *11*, 6883.
40. Cavus, N.; Zabadi, T. A comparison of open source learning management systems. *Procedia-Soc. Behav. Sci.* 2014, *143*, 521–526.
41. Buenaño-Fernández, D.; Gil, D.; Luján-Mora, S. Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability* 2019, *11*, 2833.
42. Hussain, M.; Zhu, W.; Zhang, W.; Abidi, S.M.R. Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Comput. Intell. Neurosci.* 2018, *2018*, 6347186.
43. Daghestani, L.F.; Ibrahim, L.F.; Al-Towirgi, R.S.; Salman, H.A. Adapting gamified learning systems using educational data mining techniques. *Comput. Appl. Eng. Educ.* 2020, *28*, 568–589.
44. Hofmann, T. Probabilistic Latent Semantic Indexing. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, CA USA, 15–19 August 1999; Association for Computing Machinery: New York, NY, USA, 1999; pp. 50–57.
45. Blei, D.; Ng, A.; Jordan, M. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 2003, *3*, 993–1022.
46. Nagori, R.; Gnanasekaran, A. LDA based integrated document recommendation model for e-learning systems. In Proceedings of the 2011 International Conference on Emerging Trends in

Networks and Computer Communications (ETNCC), Udaipur, India, 22–24 April 2011; pp. 230–233.
47. Zhong, J.; Zhang, S.F.; Guo, W.L.; Li, X. TFLA:A Quality Analysis Framework for User Generated Contents. *Tien Tzu Hsueh Pao/Acta Electron. Sin.* 2018, *46*, 2201–2206.
48. Yan, X.; Guo, J.; Lan, Y.; Cheng, X. A biterm topic model for short texts. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; ACM: New York, NY, USA, 2013; pp. 1445–1456.
49. Romero, C.; Ventura, S. Educational Data Mining: A Review of the State of the Art. *Syst. Man Cybern. Part C* 2010, *40*, 601–618.
50. Kotsiantis, S.; Pintelas, P. Predicting Students Marks in Hellenic Open University. In Proceedings of the Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05), Kaohsiung, Taiwan, 5–8 July 2005; Volume 2005, pp. 664–668
51. Van Rooyen, A. Distance Education Accounting Students' Perceptions of Social Media Integration. *Procedia Soc. Behav. Sci.* 2015, *176*, 444–450.
52. Samigulina, G.; Samigulina, Z. Intelligent System of Distance Education of Engineers, Based on Modern Innovative Technologies. *Procedia Soc. Behav. Sci.* 2016, *228*, 229–236.
53. Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective.Journal of Interactive Online Learning, 12(1), 126.
54. Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning.TechTrends, 59(1), 64–71. doi: https://doi.org/10.1007/s11528 -014-0822-x
55. Gedera, D., Williams, J., & Wright, N. (2015). Identifying factors influencing students' motivation and engagement in online courses. In C. Koh (Ed.). Motivation, l eadership and curriculum design (pp. 13–23). Singapore: Springerhttps://doi.org/10.1007/978-981-287-230-2_2