Original Research Paper

Framework for the Adaptive Learning of Higher Education Students in Virtual Classes in Peru Using CRISP-DM and Machine Learning

Maryori Bautista, Sebastian Alfaro and Lenis Wong

Department of Information Systems Engineering Program, Peruvian University of Applied Sciences, Lima, Peru

Corresponding Author:
Lenis Wong
Department of Information Systems Engineering Program, Peruvian University of Applied Sciences, Lima, Peru
Email: lwongpuni@gmail.com

Abstract: During the COVID-19 pandemic, virtual education played a significant role around the world. In post-pandemic Peru, higher education institutions did not entirely dismiss the online education modality. However, this virtual education system maintains a traditional teaching-learning model, where all students receive the same content material and are expected to learn in the same way; as a result, it has not been effective in meeting the individual needs of students, causing poor performance in many cases. For this reason, a framework is proposed for the adaptive learning of higher education students in virtual classes using the Cross-Industry Standard Process for Data Mining (CRISP-DM) and Machine Learning (ML) methodology in order to recommend individualized learning materials. This framework is made up of four stages: (i) Analysis of student aspects, (ii) Analysis of Learning Methodology (LM), (iii) ML development and (iv) Integration of LM and ML models. (i) evaluates the student-related factors to be considered in adapting their learning content material. (ii) Evaluate which LM is more effective in a virtual environment. In (iii), Four ML algorithms based on the CRISP-DM methodology are implemented. In (iv), The best ML model is integrated with the LM in a virtual class. Two experiments were carried out to compare the traditional teaching methodology (experiment I) and the proposed framework (experiment 2) with a sample of 68 students. The results showed that the framework was more effective in promoting progress and academic performance, obtaining an Improvement Percentage (IP) of 39.72%. This percentage was calculated by subtracting the grade average of the tests taken at the beginning and end of each experiment.

Keywords: Adaptive Learning, CRISP-DM, Machine Learning, Virtual Classes

Introduction

Adaptive learning personalizes the teaching process based on the individual needs and preferences of students using technology to overcome learning barriers (Vilela *et al.*, 2021). However, Peruvian higher education institutions still apply a traditional education system in the adoption of virtuality as a result of the COVID-19 pandemic. In other words, teachers use a teaching method in which all students learn the same topics in the same way and at the same time, dismissing the incorporation of new digital paradigms that improve academic training (Arias *et al.*, 2020). As a result, the academic performance of higher education students. who learn in different ways and at different paces, become significantly affected.

According to the World Economic Forum, Peru iranked 127th out of 137 in terms of education quality, which demonstrates that the Peruvian education system is deficient in the academic training of students (Espinoza, 2020). A study shows that 67% of the students of SENATI's industrial administration school obtained poor academic performance in Mathematics during the 2020 cycle (Vasquez Berrocal, 2020). Another research reveals that 71.4% of the students of Ricardo Palma University obtained a low academic performance in some courses during the 2020-2 cycle (Otero *et al.*, 2021).

To mitigate the problem, various proposals have been made. Lincke *et al.* (2021) analyzed evaluation records to predict learning outcomes for teachers to take action and increase the number of students passing. Singh *et al.* (2022)

recommend an individual study plan for each student and define tutoring strategies based on the student's learning style and level of knowledge through the SeisTutor system. Yang *et al.* (2021) implemented a summarized class material recommendation system to refer students to specific pages containing relevant information that they have to thoroughly study before class to improve academic performance. However, Lincke *et al.* (2021) do not evaluate relevant variables other than grades to obtain a more accurate prediction. Singh *et al.* (2022), the tool is focused on recommending actions for the teacher to support students, which does not allow their autonomous development. Lastly, Yang *et al.* (2021) do not recommend enough diversity of materials that adjust to each student.

Therefore, this study proposes a framework to personalize the learning of higher education students in a virtual environment through the adoption of Machine Learning (ML) as a tool supporting their learning process. The technology used as part of the framework will allow the recommendation of learning resources that adapt to the level of knowledge and Visual, Auditory, Read-write, and Kinesthetic (VARK) learning styles of each student. In this way, we seek to improve the quality learning experience of students and thus motivate them to continue learning and building skills. The framework proposed consists of four phases: (1) Selection of student aspects, (2) Selection of learning methodology, (3) Machine learning development, and (4) Integration of the learning methodology and machine learning model.

Various studies on adaptive learning have been found in the literature, such as the application of Artificial Intelligence (IA) algorithms (Hamim *et al.*, 2021; Lincke *et al.*, 2021; Yanes *et al.*, 2020), pedagogical methodologies applied in adaptive learning (Clark and Kaw, 2020; Clark *et al.*, 2022) and the consideration of different student aspects to individualize learning (Hariyanto and Köhler, 2020; White, 2020).

IA algorithms

Literature shows that five ML algorithms have been applied for the following purposes: Prediction of the student's performance (Evangelista, 2021; Lincke *et al.*, 2021; Qiu *et al.*, 2022; Sense *et al.*, 2021), recommendation of appropriate actions to improve the quality of courses (Hosny and Elkorany, 2022; Yanes *et al.*, 2020), recommendation of individualized learning resources (Arsovic and Stefanovic, 2020; Cheng and Wang, 2021; Ling and Chiang, 2022), prediction of the best academic engineering program for the student (Ezz and Elshenawy, 2020). Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM) have been used to predict the performance of students' evaluation, being DT and RF the most accurate algorithms, exceeding 90 % (Evangelista, 2021). On the other hand, K-means clustering, DT, LR, RF, and SVM have been the

algorithms used to recommend appropriate actions for teachers to improve the academic performance of their students. K-means obtained the best accuracy, with 93% (Hosny and Elkorany, 2022), while DT achieved the best accuracy, with 69.23% (Ling and Chiang, 2022), for the recommendation of study content material. LR, RF, and SVM algorithms were used to recommend the most suitable engineering department for each student, LR obtained the best measure of accuracy based on both the precision and the recall, with 91% (Ezz and Elshenawy, 2020).

Learning Methodologies

Three learning methodologies allowed for providing a set of tools to facilitate the teaching-learning process. Among the most used in adaptive learning were flipped classroom and spaced learning. Although both are applicable in virtual classes and are intended to improve the effectiveness of teaching-learning (Clark and Kaw, 2020; Sense *et al.*, 2021), they have different approaches.

On the one hand, flipped classroom focuses on the study of learning resources outside and inside the classroom to promote greater student participation (Clark *et al.*, 2022). On the other hand, spaced learning involves distributing learning material in small batches over time, rather than presenting it all at once, to encourage information retention (Cheng and Wang, 2021). Regarding the third methodology, the Testing effect, few studies apply it; it focuses on the repetitive study of previously learned material to remember objective information such as facts, dates, and definitions, among others, but it is not effective for learning more complex or abstract concepts such as mathematics (Sense *et al.*, 2021).

Student Aspects

The adoption of adaptive learning has considered five student aspects. The studies that considered VARK learning styles and prior knowledge agree that there was greater acceptance of the recommended content and better acquisition of new knowledge and skills (Arsovic and Stefanovic, 2020; Peng and Fu, 2022; Singh *et al.*, 2022). The reason is that these factors promoted the participation and motivation of students. Grade-based adaptive learning also led to positive results, as it provided a clear indication of the student's level of understanding of a particular topic. However, the articles Clark *et al.* (2022); White (2020) that exclusively considered this factor failed to greatly satisfy the learning experience of students (Clark and Kaw, 2020; Tavakoli *et al.*, 2022). For their part, the articles that used the speed of learning of students and interactions within a learning platform limited the ability of the system to provide an adequate recommendation, since these factors do not provide a complete picture of the learning process of students (Hamim *et al.*, 2021; Hosny and Elkorany, 2022; Ling and Chiang, 2022).

Table 1 summarizes the pros and cons of the research works mentioned in the previous paragraphs.

Table 1: Comparison of pros and cons in related work

	Related work
Pros	The algorithms used have an accuracy greater than 90%
	The recommendations promoted participation and motivation of students
	Positive results in academic performance
Cons	Studies do not evaluate relevant variables to obtain a more accurate prediction
	The aspects considered limited the system's ability to provide an adequate recommendation
	The studies are focused individually: Teacher or student, but both are not involved in a better academic result

Materials and Methods

Figure 1 describes the framework for the adaptive learning of higher education students in virtual classes using machine learning to recommend learning materials that adapt to their individual needs and preferences. The proposal consists of four phases: (1) Selection of student aspects, (2) Selection of learning methodology, (3) Machine learning development, and (4) Integration of the learning methodology and machine learning model.

Selection of Student Aspects

Table 2 shows the five aspects considered for adaptive learning according to the literature. The selection is made based on the two aspects most used and with the best results: (i) The VARK learning style (Cavanagh *et al.*, 2020) and (ii) The level of prior knowledge (Arsovic and Stefanovic, 2020).

First, the selection of VARK learning styles favors students by providing materials that match their learning preferences, which can increase their motivation and commitment to learning. Then, the level of knowledge is selected to measure the level of competence that each student has on a specific topic.

Integrating these two components will allow learning and teaching to be more personalized and also achieve a higher level of adaptability.
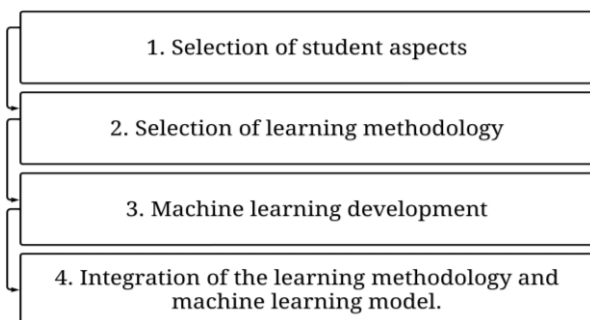

Fig. 1: Conceptualization of the framework proposed

Selection of Learning Methodology

In this stage, benchmarking is carried out to choose the most appropriate learning methodology among the three identified in the literature: Testing effect (M01), flipped classroom (M02), and spaced learning (M03). To measure these methodologies, three criteria are considered: (i) Support of an LMS (C01), which refers to the use of a digital platform as a means to manage and facilitate online teaching and learning; (ii) Participation of a teacher (C02), which means that the learning methodology is supported by the teacher to guarantee the quality and effectiveness of the teaching-learning process and (iii) Collaborative work (C03), referring to the interaction between students and teacher, which fosters an active learning collaboration and participation environment.

Then, to qualify the criteria, the Likert scale is applied with a score from 1-5, where 1 is 'totally disagree' and 5 is 'totally agree' (Pescaroli *et al.*, 2020).

Finally, Table 3 shows that the flipped classroom (M02) methodology obtained the highest score (15) compared to the other two. The assignment of this score was based on the review of literature, which gives prominence to flipped classroom methodology for promoting the use of an LMS for the delivery of learning content, for incorporating online interaction tools that allow teachers to address the individual needs of students and for enabling collaborative work between students and teachers both inside and outside the classroom. Therefore, the technique applied in this research is a flipped classroom.

Machine Learning Development

The study material predictive model adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach to guide the data analysis process in this research. This approach is based on five phases: (a) Business understanding, (b) Data understanding, (c) Data preparation, (d) Modeling, and (e) Evaluation. In addition, Google Colaboratory was used to carry out the process, as it provides an interactive development environment to work with ML. Python's Scikit-learn library was also used to build, train, and evaluate ML models.

Business Understanding

This phase grasps the problem of poor academic performance of university students in virtual classes from a business perspective. This allows the collection of accurate data to solve the problem.

The objective is to personalize learning based on the needs of each student through the prediction of material resources to improve academic performance. To achieve this, variables are used, which will be explained below.

Table 2: Aspects of students considered in adaptive learning

ID	Student aspects	Description
A01	VARK learning styles	Ways of learning are categorized into visual, auditory, read/write, and kinesthetic according to Neil Fleming (Fleming and Baume, 2006; Smyrnova-Trybulska *et al*., 2022)
A02	Level of knowledge	Previous competencies are categorized into beginner, intermediate, and advanced based on evaluation (Arsovic and Stefanovic, 2020)
A03	Speed of student learning	Pace of development of the activities proposed in an educational platform (Cavanagh *et al*., 2020)
A04	Evaluation grades	Record of grades (Lincke *et al*., 2021)
A05	Interactions in the Learning Management System (LMS)	Interaction time within the educational platform (Peng and Fu, 2022)

Table 3: Aspects of students considered in adaptive learning

	Learning methodologies		
Criteria	M01	M02	M03
C01	4	5	2
C02	2	5	2
C03	2	5	1
Total	8	15	5

Table 4: Classification of predictor variables

Test	Variables	Variable type
Quiz 1	Learning style (F1)	Psychological
Quiz 2	Grade (F2)	Cognitive
	Level of knowledge (F3)	Cognitive
	Course topic (F4)	Psychological
	Time to complete test (F5)	Cognitive

Data Understanding

In this phase, the categories of variables that are part of the prediction model are defined. The variables are identified using Keller's Attention, Relevance, Confidence, and Satisfaction (ARCS) model of motivation, which states that cognitive and psychological characteristics are required to design an individualized learning environment (Wang *et al*., 2023). Cognitive characteristics relate to academic performance and acquisition of knowledge by the student; e.g., level of knowledge, grades, among others. Psychological characteristics refer to the motivational aspects and interests of the student, such as learning styles (Arsovic and Stefanovic, 2020).

To gather information on the types of variables, 563 students of the virtual basic English course of a private university in Lima were evaluated through the VARK questionnaire (quiz 1) and a knowledge test (quiz 2).

Quiz 1 consists of 16 questions that explore a person's learning preferences (Fleming and Baume, 2006) to determine their learning style. It lasts 30 min and the latest update of 2006 has been used.

Quiz 2 consists of 39 closed questions divided into 4 blocks vocabulary, language use, grammar, and writing. This test is applied to measure the level of mastery in the course. It lasts 110 min and is prepared by the teacher.

Table 4 shows the predictor variables collected from quizzes 1 and 2.

Finally, 273 links of study material (objective variable) were collected from the virtual platform on these selected topics.

Data Preparation

Data cleaning. Once the test results of the 563 students were gathered, data cleaning and categorical variable coding were carried out. VARK questionnaire divides students based on the learning style variable: Visual, auditory, read/write, or kinesthetic. However, some records indicate that students have two, three, or four learning styles. As the learning style must be only one, these records are removed (Fleming and Baume, 2006). Regarding the knowledge test, there are records of three variables: Qualification (0-20), time to complete the test (in minutes), and level of knowledge based on the grade: (i) Beginner (0-10), (ii) Intermediate (11-16) and (iii) Advanced (17-20). There are records where exams lasted less than 15 min or the entire 110 min, but students did not solve 90% of the test. These records were deemed invalid and were tossed out. Data from tests taken twice by the same student were also dismissed, considering only the first attempt. After data cleaning, a total of 500 records were obtained.

Likewise, the change of categorical variables was carried out, from the learning styles and level of knowledge variables to numeric ones. To achieve this, the one-hot coding technique is used, in which categorical records are converted to binary (Minn, 2022) since the ML models to be tested require numeric variables as input. Study topics are converted to an ordered categorical variable and each topic is then assigned a unique integer value.

Feature selection. In this step, the least useful variable from the data set is removed: Time to complete the test, referring to the time it took the student to complete the knowledge test. Based on the evaluation records, it is determined that the time to complete the test variable is subjective and does not necessarily reflect the level of competence or understanding of students. For example, there are samples in which students obtained a higher grade (11) at the same time (100 min) compared to others who obtained a lower grade (7), which means that time is not a reliable indicator to assess student performance.
Table 5 shows the predictor variables that will finally be considered for the predictive model.

Table 5: Classification of predictor variables

ID	Features	Description	Variable type
F1	VARK learning style	Student's way of learning: Visual, auditory, read/write, and kinesthetic	Psychological
F2	Knowledge test score	Indicates the score achieved in the knowledge: 0-20	Cognitive
F3	Level of knowledge	Classifies students as beginner, intermediate, or advanced based on the knowledge test results	Cognitive
F4	Topic	The subject of study preferred by the student	Psychological
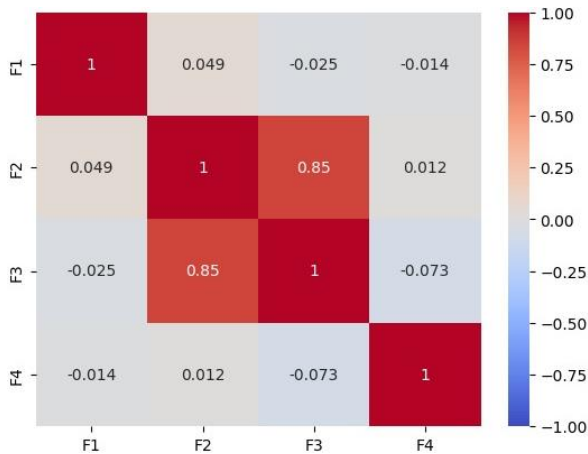
Fig. 2: Correlation matrix

In this step, the correlation analysis between the features is also carried out through the correlation matrix (Fig. 2). It is observed that grade (F1) and level of knowledge (F2) have a strong positive correlation: Pearson's correlation coefficient, 0.85, is very close to 1, which means that the greater the level of knowledge of the student, the higher the grades obtained. This can be interpreted as a validation that the level of knowledge is an important factor in academic performance.

In addition, variables are distributed (Fig. 3) to verify if the data of features present outliers and should not be considered. It is observed that all values are within the ranges initially determined. For example, the chart of the Grade variable (F1) shows that the records are between the limits of 0-20 (Fig. 3a). Based on their level of knowledge, 80 % of students are classified (F2) as Beginners (Fig. 3b). The most common learning style (F3) in students is read/write (Fig. 3c). Finally, the Output or the target variable is the learning resource.

Modeling

In this phase, the ML algorithms most used in literature for multiple classifications are applied (Arsovic and Stefanovic, 2020; Hamim *et al.*, 2021; Jiménez *et al.*, 2023; Lincke *et al.*, 2021):

Decision Tree (DT): It is a supervised ML algorithm used for classification based on one or more attributes. It has a hierarchical tree structure consisting of a root node, branches, inner nodes, and leaf nodes.

Random Forest (RF): It builds multiple decision trees that average the results of each of these individual trees. These individual trees allow predictive accuracy to increase.

Support Vector Machine (SVM): This technique is used in classification and regression problems by finding the optimal hyperplane that best classifies data points into classes or that best fits data points in regression problems.

Logistic Regression (LR): It is used only in binary or multiclass classification problems. A logistic function is applied to estimate the probability of belonging to each class.

To implement the models with these ML algorithms, the hyperparameters detailed in Table 6 are defined.

Evaluation

The performance evaluation of the algorithms is conducted based on metrics commonly used in ML: *Precision* (Eq. (1)), *Accuracy* (Eq. (2)), *Recall* (Eq. (3)), and F1-score (Eq. (4)). Micro-average Receiver Operating Characteristic (ROC) was also used:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \tag{4}$$

where:

- *True Positives* (*TP*): Learning resources correctly classified as relevant based on the needs of students
- *True Negatives* (*TN*): Learning resources correctly classified as not relevant based on the needs of students
- *False Positives* (*FP*): Learning resources incorrectly classified as relevant based on the needs of students
- *False Negatives* (*FN*): Learning resources incorrectly classified as not relevant based on the needs of students

In the classification of multiclass models, *precision* or *TN* rate is defined as the number of cases in which the model correctly predicts a specific class as positive. Accuracy measures the proportion of cases correctly classified about the total number of samples. *Recall* or *TP* rate represents the number of cases correctly classified as positive for a specific class. *F1-Score* evaluates the performance of the model in terms of the correct classification of positive and negative cases (Arsovic and Stefanovic, 2020).

Table 6: Classification of predictor variables

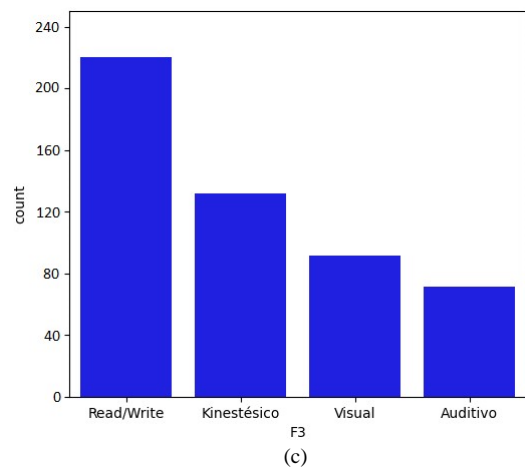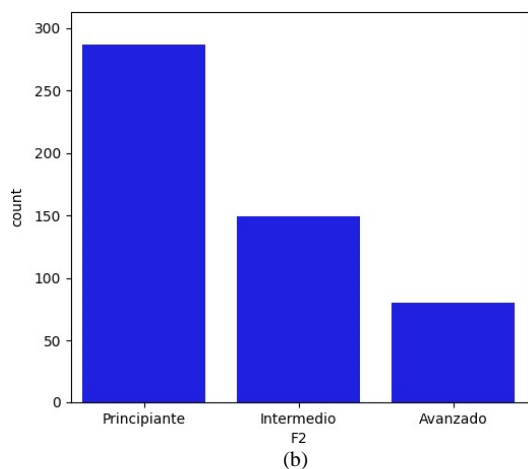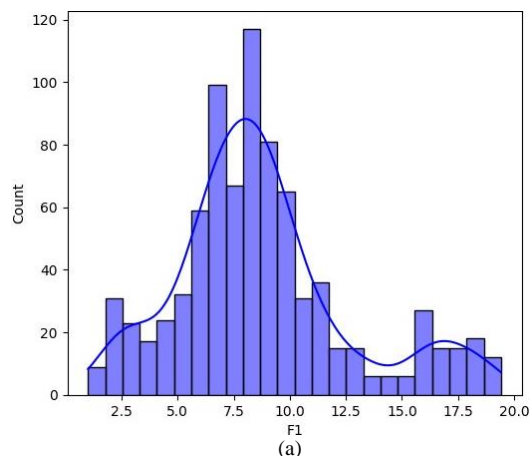| Algorithms	Hyperparameter applied
DT	Maximum number of levels of 10 minimum number of samples required to perform a division of 2
RF	Maximum tree level of 10, 2 samples at one node to consider a further split 100 estimators
SVM	Regularization C in 100
LR	Regularization C in 100

(a)

(b)

(c)

Fig. 3: Distribution of the variables: Grade- F1(a), level of knowledge- F2 (b) and VARK learning style-F3 (c)

Additionally, the graphical representation of the micro-average ROC curve summarizes the overall performance of multiclass classification models by combining all classes into a single curve. To obtain this curve, dataset instances are considered. and the TP and FP rates are calculated for each class to then calculate a weighted average and obtain the micro-average.

Regarding the chart, the model getting too close to the midpoint indicates that all classes are being classified in a similar way, without adequately distinguishing between them. Area Under the Curve (AUC) is a metric derived from the micro-average ROC curve that represents the probability that the model will correctly classify a randomly chosen instance, where the correct classification is defined as the instance belonging to the correct class according to the actual label (Lincke *et al.*, 2021). AUC values are in the range of 0-1. A value closer to 1 indicates a better model performance in multiclass classification.

Integration of Learning Methodology and the Recommendations of the ML Model

In this stage, the selected learning methodology (flipped classroom) is integrated (Clark and Kaw, 2020) with the model with the best performance (RF). Figure 4 shows the diagram of the integration of the learning methodology and ML model; it shows the use of the recommended study materials in the three phases of the learning methodology: (1) Before class, (2) During class, and (3) after class.

The framework was validated with 68 students of the basic English course who participated in two different experiments to measure the level of learning achieved: (a) Traditional teaching-learning process and (b) Teaching-learning process using the framework in a class session. Both experiments were carried out throughout two weeks of class with the same teacher and included two groups each. Also, it considered the same English topics, such as the verb 'to be' and past simple.

In the first week of May, the first groups of each experiment participated with 17 students each; in the fourth week of May, the second groups of each experiment participated also with 17 students each. Table 7 shows the details of the distribution of students per experiment.

At the end of the experiments, the following metrics are evaluated: Grade Average (FA), Grade Mode (GM), and Percentage of Improvement (IP).
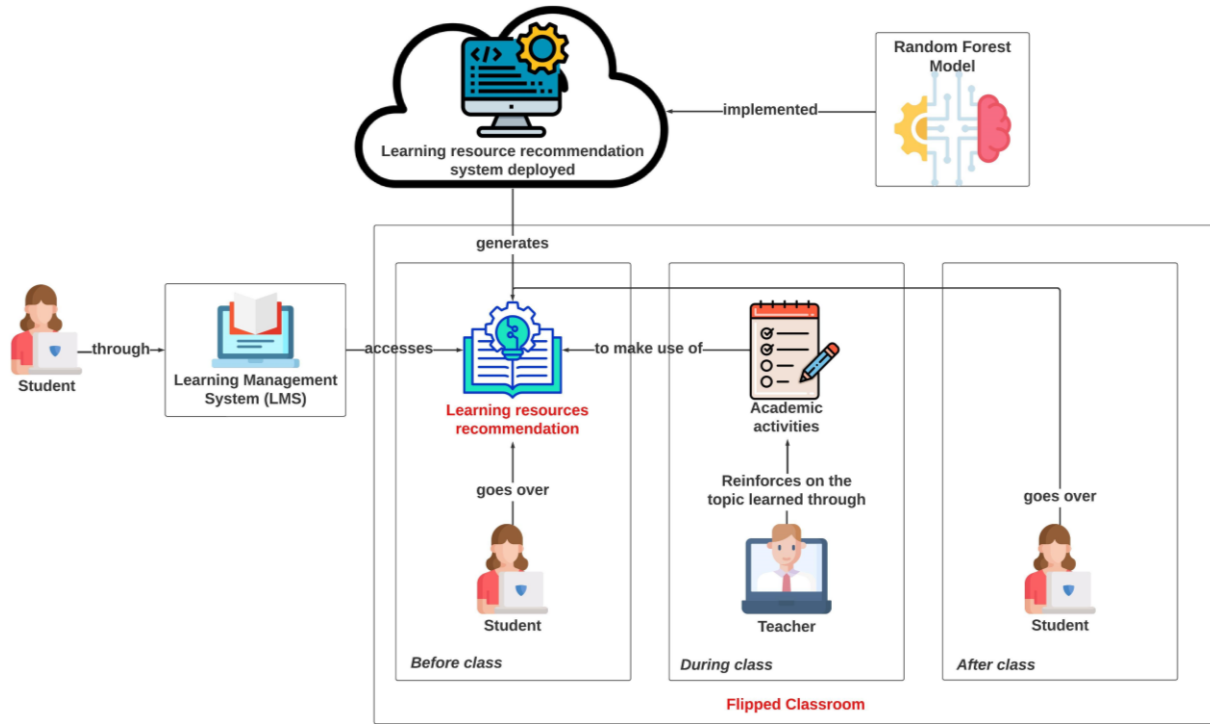
Fig. 4: Diagram of the integration of the learning methodology and the recommendations of the RF model

Table 7: Summary of the two experiments carried out to validate the level of learning

Experiment 1: Traditional learning			
Participants	Date	Topics	Metrics
Group 1	1st week of May	Topic 1: Verb 'to be'	Grade average, grade mode and percentage improvement
Group 2	4th week of May	Topic 2: Past simple: Regular and irregular verbs	
Experiment 2: Learning with the framework proposed			
Group 1	1st week of May	Topic 1: Verb 'to be'	Grade average, grade mode and percentage improvement
Group 2	4th week of May	Topic 2: Past simple: Regular and irregular verbs	

Experiment I: Traditional Teaching-Learning Process

Experiment 1 consisted of developing the new topics corresponding to the two weeks of class under traditional learning. This experiment involved the following steps: (i) The topics were studied for the first time in the synchronous session. (ii) All the students developed the same activities in their LMS; e.g., homework to go over the topic. (iii) At the end of the week, the teacher prepared and sent a 30-minute evaluation of the two topics learned to measure the level of learning achieved. Fig. 5 shows the learning process of each group of experiment I.

Table 8 shows the results of the initial knowledge test taken by the two groups of experiment I to be compared with the results of the final knowledge test taken at the end of the experiment.
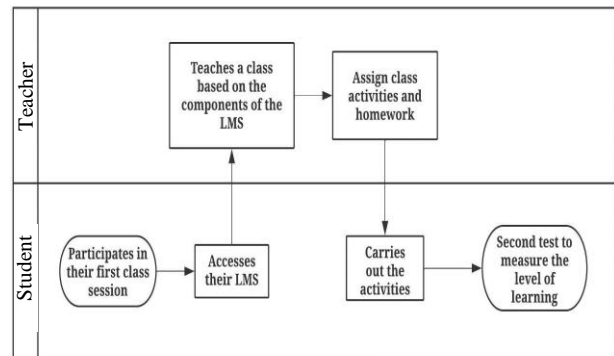

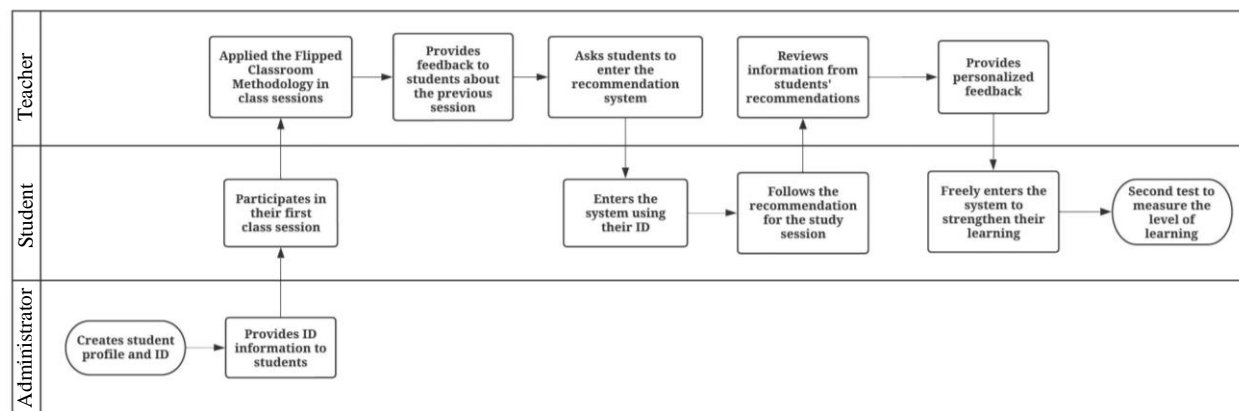**Fig. 5:** Traditional teaching-learning process in a virtual class session

Fig. 6: Traditional teaching-learning process in a virtual class session

Table 8: Results of initial knowledge test of students from experiment I

Statistics	Group 1	Group 2
Average grade	8.18	8.18
Minimum grade	7.14	7.13
Maximum grade	9.45	9.77

Experiment II: Teaching-Learning Process Using the Framework

Experiment II consisted of developing new topics corresponding to the two weeks of class and applying the framework. This experiment involved the following steps: (i) Students developed class sessions under the Flipped Classroom methodology, relying on the material recommendation system before their first synchronous session with the teacher. (ii) In class, they received individual feedback from the teacher based on the resources suggested by the system. (iii) At the end of the session, students had the possibility of accessing the system and obtaining more content recommendations to reinforce the knowledge learned in class. (iv) At the end of the week, the teacher prepared and sent a 30 min evaluation of the two topics learned to measure the level of learning achieved. Figure 6 shows the learning process of each group in experiment II. In the three steps, the students accessed Moodle's LMS to select a topic and view recommendations through a new section created in each classroom.

Table 9 shows the results of the initial knowledge test taken by the two groups of experiment II to compare with the results of the final test at the end of the experiment.

Results and Discussion

ML Algorithms

To calculate the performance of the selected models, 20% of the dataset is used, that is, a sample of 100 instances, as 80% of the dataset is destined for training. As a result of training the four selected algorithms, the micro-average ROC curve is obtained (Fig. 7). RF is the algorithm with the best AUC value, with 0.9872. LR and SVM also show strong performance in multiclass classification, but they are below the RF model, with 0.9119 and 0.9052, respectively. DT presents a value of 0.8189, being the closest to 0.5, which means that this algorithm tends to classify on a random basis. Figure 8 shows the micro-average ROC curve of each algorithm.

The ML algorithms defined in the modeling phase are compared under the previously defined evaluation metrics (Table 10). The table shows that the algorithms based on tree structures, such as RF and DT, obtained a precision of 98% for the recommendation of learning resources. However, the RF stands out when evaluated in its accuracy and F1-Score. RF was the algorithm with the highest accuracy rate of 92%. Compared to the results obtained by Ling and Chiang (2022), the metric obtained from the RF algorithm showed a better accuracy rate compared to 69.23% of the study.

On the other hand, the algorithm with the lowest accuracy is SVM, with 3%. As for precision, the LR algorithm is the one with the lowest results.

Therefore, based on the evaluation of the four ML algorithms, the RF algorithm obtained the best results. It is the best performing in classifying and recommending learning resources for different classes.

Experimentation

The results of the level of learning achieved by the groups of students were determined by the grade average of the test taken at the end of the experiment (*FA*) and the Improvement Percentage (*IP*) calculated with Eq. (5):

$$IP = \frac{F_A - I_A}{I_A} \times 100 \qquad (5)$$

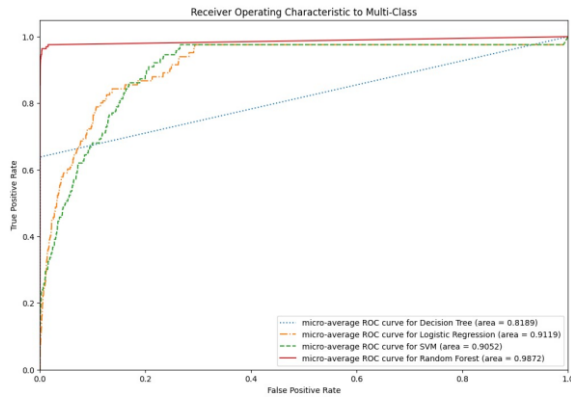where:
FA : Grade average of the final test
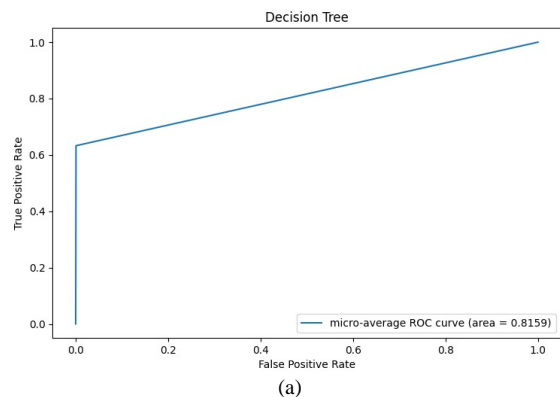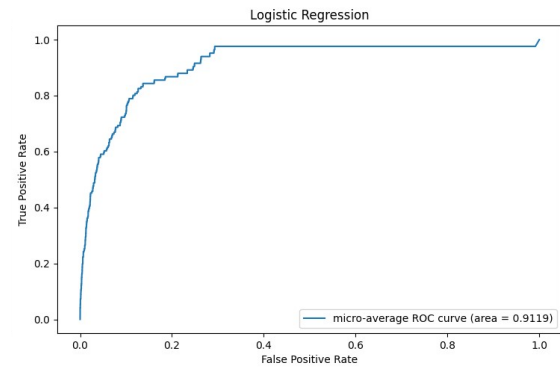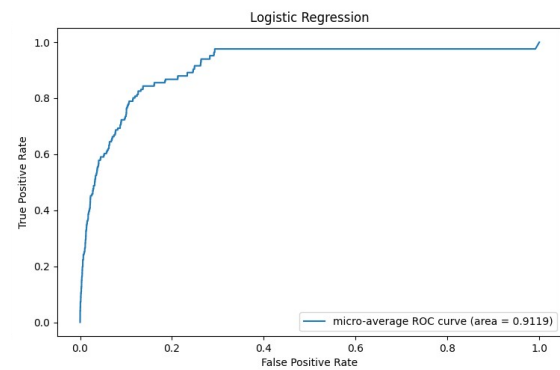*IA* : Grade average of the initial test

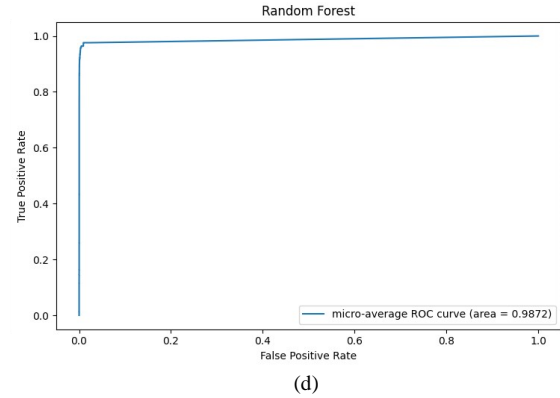**Fig. 7:** Micro-average ROC curve comparison

(a)

(b)

(c)

(d)

Fig. 8: Micro-average ROC curves of algorithms used in model development: (a) DT; (b) LR; (c) SVM; (d) RF

Table 9: Results of initial knowledge test of students of experiment II

Statistics	Group 1	Group 2
Average grade	8.69	8.46
Minimum note	6.63	7.07
Maximum note	9.96	9.63

Table 10: ML metrics obtained from the training of the algorithms

ML Algorithms	Metrics			
	Precision	F1-score	Recall	Accuracy
DT	0.98	0.91	0.92	0.91
LR	0.86	0.17	0.22	0.22
RF	0.98	0.92	0.92	0.92
SVM	0.96	0.01	0.03	0.03

Table 11 shows the results obtained from the two groups of students under traditional learning (experiment I). Group 1 obtained the highest average in the final test taken at the end of the experiment (FA), with 12. Based on these grades, the equation to calculate the IP was applied, with Group 1 obtaining the highest learning improvement percentage, 48.19%. Group 1 also obtained the highest Grade Mode (GM), with the most frequent grade being 14. Finally, averaging both groups, FA, GM, and IP obtained 12.06, 12, and 47.44%, respectively.

The charts below represent the evolution of grades of both Group 1 (Fig. 9a) and Group 2 (Fig. 9b) of experiment I. They compare the grades obtained in the knowledge test before the experimentation and those obtained in the test carried out by the teacher at the end of the experiment, evidencing slight progress in their learning, as reflected in Table 11.

Table 12 shows the level of learning results obtained by the two groups of students who used the framework (experiment II). When averaging the final tests of students, group 2 obtained the highest grade, 16.24. When estimating the IP, group 2 obtained the highest learning improvement percentage, 91.81%. For its part, the GM of both group 1 and group 2 was 18. Finally, the FA, GM, and IP averages of both groups were 16, 18, and 86.65%, respectively.

530

**Table 11:** Results of the level of learning achieved at the end of experiment I

Group	IA	FA	GM	IP %
1	8.18	12.12	14	48.19
2	8.18	12:00	10	46.68
Average	8.18	12.06	12	47.44

Table 12: Results of the level of learning achieved at the end of experiment II

Group	IA	FA	GM	IP %
1	8.69	15.76	18	81.49
2	8.46	16.24	18	91.81
Average	8.58	16.00	18	86.65

Table 13: Comparison of the level of learning achieved in experiments I and II

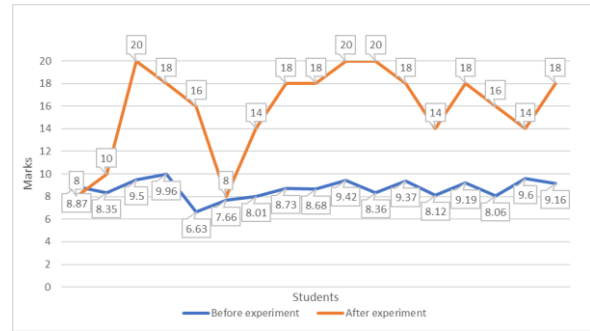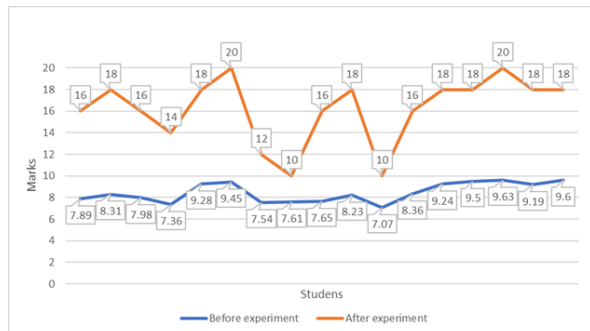| Metric	Experiment I			Experiment II		
	Group 1	Group 2	Average	Group 1	Group 2	Average
FA	12.12	12:00	12.06	15.76	16.24	16.00
GM	14.00	10.00	12.00	18.00	18.00	18.00
IP	48.19%	46.68%	47.44%	81.49%	91.81%	86.65%
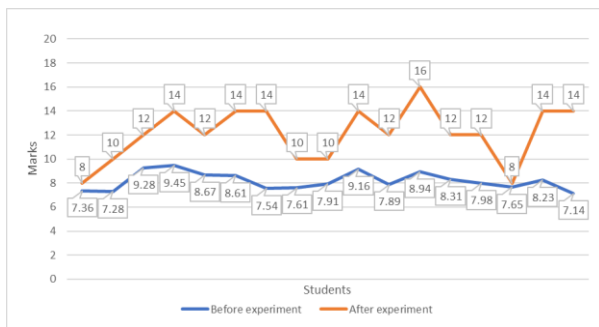
(a)
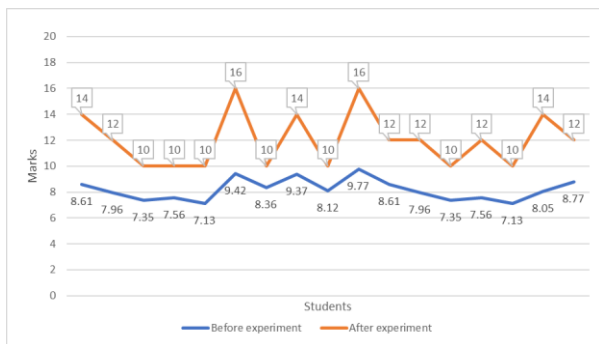
(b)

Fig. 9: Evaluation results of experiment I. Group 1 (a), group 2 (b)

The charts below represent the evolution of the grades of both Group 1 (Fig. 10a) and Group 2 (Fig. 10b) of experiment II. They show the grades obtained in the knowledge test b before experimentation compared to those obtained in the test carried out by the teacher at the end of the experiment, evidencing significant progress in their learning, as reflected in Table 12.

(a)

(b)

Fig. 10: Evaluation results of Experiment II. Group 1 (a), group 2 (b)

The results showed that the FA, GM, and IP averages in experiment I were 12.06, 12 and 47.44%, respectively (Table 11), while FA, GM, and IP averages in experiment II were 16, 18 and 86.65%, respectively (Table 12), which means that IP increased by 39.21 percentage points. It represents an improvement in academic performance.

Finally, Table 13 shows the comparison of the results of the metrics evaluated in experiments I and II. When calculating the average of groups 1 and 2 of each experiment, the following results are obtained: (i) The FA of the students who were part of Experiment I was 12, a lower performance level compared to the students of experiment II, whose FA was 16; demonstrates that those who used the framework achieved a better understanding and mastery of the topics taught. (ii) The GM of 12 was obtained by averaging the values of the groups of experiment I proving that the methodology used by experiment II was more effective in terms of helping students achieve better grades since the most frequent grade (18) was higher. (iii) The IP of students of experiment I was 47.44%, while the students of experiment II obtained 86.65%; the difference of 39.21% points shows that the teaching methodology based on the framework was more effective in promoting progress and academic performance, compared to those who followed a traditional methodology.

Conclusion

This study proposed a framework to adapt learning to the knowledge and learning style of each student and improve their academic performance through the adoption of ML.

To achieve this, four ML algorithms were compared: DT, LR, SVM, and RF. The algorithm with the highest accuracy rate was RF, with 92%.

The proposal was validated through two experiments with students of the basic English course at a private university in Lima. The first experiment consisted of developing virtual class sessions for two weeks in the traditional way with the participation of two groups of students on different days. The second experiment consisted of developing virtual class sessions for two weeks applying the proposed framework with the participation of two groups of students also on different days.

The average GM, FA, and IP of the groups of experiment I was 12.06, 12, and 47.44%, respectively, compared to 16, 18, and 86.65%, respectively, in experiment II, which reveals that the proposal led to better academic performance. These findings support the usefulness and effectiveness of the adoption of the framework as a tool to improve the quality of learning in higher education environments.

In the future, it is recommended to adapt the framework to a blended education environment. For this, new variables and technologies should be explored to increasingly improve the experience of students. For example, the present study did not consider the students' memory retention capacity, which is also an important cognitive variable that influences the students' learning process. Key considerations include student engagement, motivation, and collaboration, along with integrating emerging technologies like virtual reality or gamification. These adaptations not only benefit students but also have broader implications for education.

Acknowledgment

We would like to thank the students who took part in this research.

Funding Information

Author's Contributions

Maryori Bautista and Sebastian Alfaro: Literature analysis, data collection, modeling, evaluation of machine learning models, implementation of the selected model, experimentation and analysis of results. Manuscript written.

Lenis Wong: Study supervision, result analysis, manuscript reviewed and discussion.

Ethics

The article is authentic and contains unpublished material. The corresponding author affirms that no ethical concerns exist and all authors have read and endorsed the article.

References

Arias, E., Escamilla, J., López, A., & Peña, L. (2020). Cómo perciben los docentes la preparación digital de la Educación Superior en América Latina. *México: Observatorio de Innovación Educativa-Instituto Tecnológico de Monterrey y el BID. Recuperado de:*

Arsovic, B., & Stefanovic, N. (2020). E-learning based on the adaptive learning model: Case study in Serbia. *Sādhanā*, *45*(1), 266. https://doi.org/10.1007/S12046-020-01499-8/METRICS

Cavanagh, T., Chen, B., Lahcen, R. A. M., & Paradiso, J. R. (2020). Constructing a design framework and pedagogical approach for adaptive learning in higher education: A practitioner's perspective. *International Review of Research in Open and Distributed Learning*, *21*(1), 173-197. https://doi.org/10.19173/IRRODL.V21I1.4557

Cheng, J., & Wang, H. (2021). Adaptive algorithm recommendation and application of learning resources in English fragmented reading. *Complexity*, *2021*, 1-11. https://doi.org/10.1155/2021/5592534

Clark, R. M., & Kaw, A. (2020). Adaptive learning in a numerical methods course for engineers: Evaluation in blended and flipped classrooms. *Computer Applications in Engineering Education*, *28*(1), 62-79. https://doi.org/10.1002/CAE.22175

Clark, R. M., Kaw, A. K., & Braga Gomes, R. (2022). Adaptive learning: Helpful to the flipped classroom in the online environment of COVID?. *Computer Applications in Engineering Education*, *30*(2), 517-531. https://doi.org/10.1002/CAE.22470

Espinoza, F. (2020). Educación en el Perú:¿ Cuál es la situación de la Educación Básica Regular. *Chiqaqnews, Facultad de Letras y Ciencias Humanas-UNMSM*.

Evangelista, E. (2021). A hybrid machine learning framework for predicting students' performance in virtual learning environment. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(24), 255-272. https://doi.org/10.3991/IJET.V16I24.26151

Ezz, M., & Elshenawy, A. (2020). Adaptive recommendation system using machine learning algorithms for predicting student's best academic program. *Education and Information Technologies*, *25*, 2733-2746. https://doi.org/10.1007/S10639-019-10049-7/METRICS

Fleming, N., & Baume, D. (2006). Learning Styles Again: VARKing up the right tree!. *Educational Developments*, *7*(4), 4. https://vark-learn.com/wp-content/uploads/2014/08/Educational-Developments.pdf

Hamim, T., Benabbou, F., & Sael, N. (2021). Survey of machine learning techniques for student profile modeling. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(4), 136-151. https://doi.org/10.3991/IJET.V16I04.18643

Hariyanto, D., & Köhler, T. (2020). A web-based adaptive e-learning application for engineering students: An expert-based evaluation. *International Journal of Engineering Pedagogy*, *10*(2), 60-71. https://doi.org/10.3991/IJEP.V10I2.11834

Hosny, K., & Elkorany, A. (2022). Applying adaptive learning by integrating semantic and machine learning in proposing student assessment model. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*(2), 2014-2025. https://doi.org/10.11591/IJECE.V12I2.PP2014-2025

Jiménez, O., Jesús, A., & Wong, L. (2023, May). Model for the Prediction of Dropout in Higher Education in Peru applying Machine Learning Algorithms: Random Forest, Decision Tree, Neural Network and Support Vector Machine. In 2023 33rd Conference of Open Innovations Association (*FRUCT)* (pp. 116-124). IEEE. https://doi.org/10.23919/FRUCT58615.2023.10143068

Lincke, A., Jansen, M., Milrad, M., & Berge, E. (2021). The performance of some machine learning approaches and a rich context model in student answer prediction. *Research and Practice in Technology Enhanced Learning*, *16*(1), 1-16. https://doi.org/10.1186/S41039-021-00159-7/TABLES/5

Ling, H. C., & Chiang, H. S. (2022). Learning performance in adaptive learning systems: A case study of web programming learning recommendations. *Frontiers in Psychology*, *13*, 770637. https://doi.org/10.3389/FPSYG.2022.770637/BIBTEX

Vasquez Berrocal, Y. M. (2020). Desempeño docente y rendimiento académico de la asignatura de matemáticas en los estudiantes de estudios generales de la Escuela de Administración Industrial de una institución educativa tecnológica. https://hdl.handle.net/20.500.12866/8445

Minn, S. (2022). AI-assisted knowledge assessment techniques for adaptive learning environments. *Computers and Education: Artificial Intelligence*, *3*, 100050. https://doi.org/10.1016/J.CAEAI.2022.100050

Otero, J. M., Indacochea, S., Soto, A., & De la Cruz, J. (2021). Factores asociados a bajo rendimiento académico en estudiantes de medicina humana de la Universidad Ricardo Palma durante el período agosto-diciembre del 2020. *Revista De La Sociedad Peruana De Medicina Interna*, *34*(3), 93-99. https://doi.org/10.36393/spmi.v34i3.608

Peng, P., & Fu, W. (2022). A pattern recognition method of personalized adaptive learning in online education. *Mobile Networks and Applications*, *27*(3), 1186-1198. https://doi.org/10.1007/S11036-022-01942-6/METRICS

Pescaroli, G., Velazquez, O., Alcántara-Ayala, I., Galasso, C., Kostkova, P., & Alexander, D. (2020). A likert scale-based model for benchmarking operational capacity, organizational resilience and disaster risk reduction. *International Journal of Disaster Risk Science*, *11*, 404-409. https://doi.org/10.1007/S13753-020-00276-9/TABLES/1

Qiu, F., Zhu, L., Zhang, G., Sheng, X., Ye, M., Xiang, Q., & Chen, P. K. (2022). E-learning performance prediction: Mining the feature space of effective learning behavior. *Entropy*, *24*(5), 722. https://doi.org/10.3390/E24050722

Sense, F., van der Velde, M., & van Rijn, H. (2021). Predicting university students' exam performance using a model-based adaptive fact-learning system. *Journal of Learning Analytics*, *8*(3), 155-169. https://doi.org/10.18608/jla.2021.6590

Singh, N., Gunjan, VK, Mishra, AK, Mishra, RK, & Nawaz, N. (2022). SeisTutor: A custom-tailored intelligent tutoring system and sustainable education. *Sustainability,* *14*(7), 4167. https://doi.org/10.3390/SU14074167

Smyrnova-Trybulska, E., Morze, N., & Varchenko-Trotsenko, L. (2022). Adaptive learning in university students' opinions: Cross-border research. *Education and Information Technologies*, *27*(5), 6787-6818. https://doi.org/10.1007/S10639-021-10830-7/FIGURES/8

Tavakoli, M., Faraji, A., Vrolijk, J., Molavi, M., Mol, S. T., & Kismihók, G. (2022). An AI-based open recommender system for personalized labor market driven education. *Advanced Engineering Informatics*, *52*, 101508. https://doi.org/10.1016/J.AEI.2021.101508

Vilela, P., Sánchez, J., & Chau, C. (2021). Desafíos de la educación superior en el Perú durante la pandemia por la COVID-19. *Desde el Sur*, *13*(2). https://doi.org/10.21142/DES-1302-2021-0016

Wang, S., Christensen, C., Cui, W., Tong, R., Yarnall, L., Shear, L., & Feng, M. (2023). When adaptive learning is effective learning: Comparison of an adaptive learning system to teacher-led instruction. *Interactive Learning Environments*, *31*(2), 793-803. https://doi.org/10.1080/10494820.2020.1808794

White, G. (2020). Adaptive Learning Technology Relationship with Student Learning Outcomes. *Journal of Information Technology Education: Research, 19,* 113-130. https://doi.org/10.28945/4526

Yanes, N., Mostafa, A. M., Ezz, M., & Almuayqil, S. N. (2020). A machine learning-based recommender system for improving students learning experiences. *IEEE Access*, *8*, 201218-201235. https://doi.org/10.1109/ACCESS.2020.3036336

Yang, C. C., Chen, I. Y., Akçapınar, G., Flanagan, B., & Ogata, H. (2021). Using a summarized lecture material recommendation system to enhance students' preclass preparation in a flipped classroom. *Educational Technology and Society*, *24*(2), 1-13. http://hdl.handle.net/2433/263123