Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 13 No. 7 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijep.v13i7.40075

PAPER

# Towards an Adaptive Learning Model using Optimal Learning Paths to Prevent MOOC Dropout

El Miloud Smaili, Mohamed Daoudi, Ilham Oumaira, Salma Azzouzi, Moulay El Hassan Charaf(✉)

Ibn Tofail University, Kenitra, Morocco

my.charaf@uit.ac.ma

**ABSTRACT**

Currently, massive open online courses (MOOCs) are experiencing major developments and are becoming increasingly popular in distance learning programs. The goal is to break down inequalities and disseminate knowledge to everyone by creating a space for exchange and interaction. Despite the improvements to this educational model, MOOCs still have low retention rates, which can be attributed to a variety of factors, including learners' heterogeneity. The paper aims to address the issue of low retention rates in MOOCs by introducing an innovative prediction model that provides the best (optimal) learning path for at-risk learners. For this purpose, learners at risk of dropping out are identified, and their courses are adapted to meet their needs and skills. A case study is presented to validate the effectiveness of our approach using classification algorithms for prediction and the ant colony optimization (ACO) algorithm to optimize learners' paths.

**KEYWORDS**

adaptive learning, massive open online courses (MOOCs), ant colony optimization (ACO), PCA, drop out

## 1 INTRODUCTION

Over the past few years, there has been an increase in the number of massive open online course (MOOC) platforms providing free and open online courses to a broad audience. Particularly during the COVID-19 pandemic, when online learning became the only mode of learning around the globe, many universities started using MOOCs to reduce the obstacles, such as sanitary, financial, geographical, etc., associated with traditional learning.

Today, MOOCs have attracted quite a bit of interest from both institutional educators and online learners, as training is often provided through videos or electronic documents. As such, technological resources and tools previously unknown to students and instructors are integrated into the education environment, transforming the education landscape and providing a great opportunity to collect massive amounts of data related to students' educational achievements.

Smaili, E.M., Daoudi, M., Oumaira, I., Azzouzi, S., Charaf, M.E.H. (2023). Towards an Adaptive Learning Model using Optimal Learning Paths to Prevent MOOC Dropout. *International Journal of Engineering Pedagogy (iJEP)*, 13(7), pp. 128–144. https://doi.org/10.3991/ijep.v13i7.40075

Article submitted 2023-04-01. Revision uploaded 2023-08-18. Final acceptance 2023-08-19.

Despite their success, MOOCs experience several barriers, including learners' heterogeneity in terms of cultural and geographical backgrounds, knowledge levels, and learning habits. In fact, students who participate in MOOCs typically have varied and evolving profiles and expectations, which might lead to a very low completion rate of courses. In this context, the high dropout rate in MOOCs has drawn the attention of many professors and researchers in the field, as it highlights the importance of this learning model to the educational landscape. Therefore, we provide in this study an innovative approach to prevent MOOC dropout and to improve courses' completion.

There are a number of solutions proposed in this context, including adapting the content of MOOCs to meet the various needs and objectives of participants. However, most adaptive learning approaches recommend courses rather than learning paths.

As part of our innovative approach, we provide learners with paths to follow based on neighboring learner profiles with similar skills. More precisely, we implement our adaptive system by considering the characteristics and requirements of the learners. Based on the information provided by participants during enrollment, a profile of each learner will be generated, and their interactions with the learning environment will be used to determine their goals and preferences. In this regard, the learner's profile will play an important role in the personalization of learning paths since it allows the system to determine the most suitable courses for each student. In this process, the first stage will be to identify at-risk learners via a classification algorithm. The task can be accomplished using a variety of algorithms. For instance, a decision tree classifier predicts future observations based on a corpus of previously labeled observations. In contrast, the adaptive boosting algorithm learns from previous errors and makes a weighted prediction. Furthermore, the gradient boosting algorithm alters the instance weights with every interaction to adjust the new predictor according to the residual errors of a previous prediction. Then, in the second stage, we use the ant colony optimization (ACO) algorithm to identify the appropriate courses for each learner and thus create a pathway that is most suitable for its needs and requirements. To summarize, our model attempts to answer two fundamental research questions: How can course adaptation as well as learning path selection boost learner retention and improve course completion rates? And what are the key factors to take into account when adapting MOOC courses to meet the needs and preferences of learners according to their level of knowledge and objectives?

The remainder of the paper is organized as follows: Section 2 summarizes some preliminary concepts. The problem statement and a review of some related works are discussed in Sections 3 and 4, respectively. In Section 5, we describe the materials and explain our adaptive learning methodology. Afterwards, we present in Section 6 the major findings of our case study to optimize the model and evaluate the effectiveness of our approach. A conclusion and future work are given in the last section.

## 2 PRELIMINARIES

### 2.1 MOOCs vs. adaptive learning

Massive open online courses are open training courses available to all, which can lead to certification across the world, regardless of location [1]. In MOOCs, students can access texts, videos, forums, and online questionnaires. Furthermore, adaptive learning can indeed be relevant in addressing students' heterogeneity and dropout issues in MOOCs. Adaptive learning can be viewed as a learning method in which one or more characteristics of the learning environment adapt to the learner.

The main idea is to meet the individual needs of each learner during the learning process. Therefore, it is necessary to properly identify the profile of each learner and adapt the program accordingly, making the learner an active participant in the learning process. This is done by providing learners with tailored recommendations for educational resources, making achieving their goals easier and more efficient [2].

### 2.2 Clustering vs. classification

In data science, a cluster is a collection of data points that are grouped based on their similarity to each other. The K-means clustering algorithm is one of the most popular unsupervised clustering algorithms. The K-means algorithm divides a dataset into K distinct clusters based on the degree of similarity between the observations. In other words, the K-means algorithm identifies K-centroids and then allocates each data point to the closest cluster while keeping the centroids as small as possible [3].

As for classification, it requires the application of machine learning algorithms that learn how to assign class labels to examples in a given problem domain. In a classification problem, we attempt to classify an object into different classes, i.e., to predict the value of a discrete variable that has a finite number of values [4].

### 2.3 Evaluation metrics

In order to evaluate a machine learning model, several performance metrics are available. There is a great deal of variation in how machine learning algorithms can be measured and compared, depending on the metric used. We quote in this article the following metrics [5]:

**Precision recall and F1 score:** These metrics are calculated from a confusion matrix since they are based on true and false classifications. In classification, recall is also referred to as true positive rate or sensitivity, and precision is referred to as positive predictive value.

**AUC-ROC curve:** In classification problems, the AUC-ROC curve provides a measure of performance at a variety of threshold settings. It is relevant to note that ROC is a probability curve and AUC is a measure of separability. As a result, it is an indication of how well the model can differentiate between different classes. In general, the higher the AUC, the better the model will be able to predict classes.

## 3 PROBLEM STATEMENT

In recent years, MOOCs have become increasingly popular in the field of education. One of the major challenges facing MOOCs, however, is the high rate of dropouts before completing the course. In fact, the geographical dispersion of learners enrolled in MOOCs with various backgrounds leads to differences in their basic knowledge, needs, expectations, and learning habits.

Due to these different learning profiles, adaptive learning is becoming one of the most popular methods for meeting participants' learning needs and objectives. However, most adaptive learning approaches recommend courses instead of paths. In this study, we suggest providing learners with paths to follow based on neighboring profiles with similar skills. Moreover, for optimum results, we suggest that learners at risk be identified first before appropriate learning paths are generated. Through the adaptation of content to each learner's specific needs, we are able to ensure that our solution offers promising results and significantly increases the success rate of MOOCs.

## 4 LITERATURE REVIEW

In recent years, a wide range of studies have been conducted on MOOCs. A major focus of research in MOOCs is the problem of dropout rates. Accordingly, the authors in [6] identified variables that impact student performance during the disorientation of the educational process resulting from the COVID-19 pandemic. Similarly, another work [7] discusses how to ensure that distance education learners experience a smooth learning process. The authors discuss how to evaluate the acquired knowledge and skills in light of current events, particularly in uncertain circumstances such as the COVID-19 pandemic. In [8], the authors suggest a powerful remediation approach to enhance collaboration between students and improve knowledge and skills.

Despite the numerous studies on dropout, most of the research is concerned with: (i) predicting learners at risk without suggesting strategies for overcoming the high dropout rate; or (ii) recommending courses to all students rather than those at risk. In fact, the authors in [9] propose a generalized linear model that was trained on 70% of participants to predict grades for a further 30% based on newly discovered factors that can influence students' academic achievements. Furthermore, the authors in [10] present and discuss recommendations for improving student engagement in large-scale open learning. In this context, a comparative analysis of a variety of recommender systems within the context of MOOCs is provided. Next, the authors suggest a recommender system that provides learners with appropriate personalized learning activities as part of the learning process in order to keep them motivated and satisfied. By using learners' behavior data, the authors of [11] propose a pipeline model named CLSA for predicting dropout rates. As part of its design, CLSA utilizes a convolutional neural network to generate a high-dimensional vector and feeds this vector to a long-short-term memory network. When tested on the KDD CUP 2015 dataset, the model achieved 87.6% accuracy and an F1-score of 86.9%.

Another work [12] aims to provide learners with personalized learning paths using a collaborative optimization algorithm that combines ant colony optimization and genetic algorithms. In order to construct an optimal solution, the proposed algorithm takes advantage of the stochastic nature of ant colony optimization and the exploration characteristics of genetic algorithms. Furthermore, using a classification algorithm, namely the naive Bayesian algorithm, the work [13] proposes an approach for adapting MOOCs. Learners can follow the learning materials of their customized course to achieve specific goals instead of taking difficult or easy courses. Moreover, the article [14] proposes a learning model using ant colony algorithms. As part of the algorithm model, personalized information is incorporated, including educational background and learning style. Thus, teachers are able to create customized lessons and provide appropriate learning materials. Likewise, the authors in [15] propose an intelligent adaptive system based on artificial intelligence with the main objective of providing personalized learning environments tailored to the learners' needs.

Aside from that, the article [16] combines prediction and course personalization by exploring historical data on the MOOC Canvas Network platform. The authors also preprocessed that data to recommend the top ten most effective online courses for learners.

In the same way, we recommend offering courses only to learners at risk of dropping out and customizing learning paths accordingly. The paper can be a continuation of previous research [17] [18] [19] [20] where some solutions are provided to reduce the high drop-out rate in MOOCs and to enhance adaptive learning systems using some optimization algorithms.

# 5 ADAPTIVE LEARNING MODEL

## 5.1 Overview

The purpose of this paper is to reduce the high dropout rate through the use of personalization methods. Based on the traces left by learners during their interactions with the learning environment, we tailor the course to meet the needs of each learner. Thus, the study attempts to derive information about learners' skills and behaviors in order to facilitate the creation of their profiles. The profiles of at-risk learners play an important role in identifying the optimal paths for learners to follow using the ant colony optimization approach. Figure 1 illustrates the steps involved in creating our adaptive model.

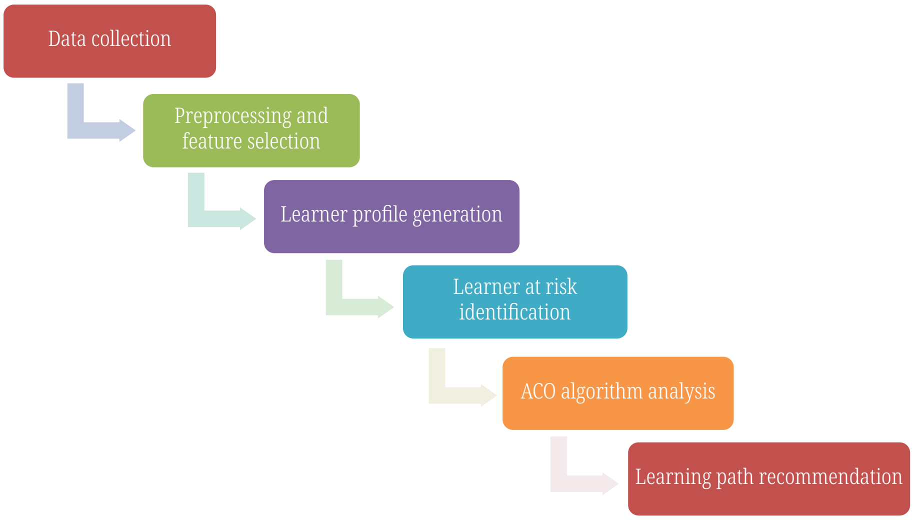



**Fig. 1.** Suggested adaptive learning model

## 5.2 Data processing

**Data collection.** The collection of data in MOOCs is crucial for making informed decisions regarding the retention rate of learners. Generally, participants in a MOOC are characterized by a profile created from the learner's personal information. This includes their identifier, age, geographical location, level of education, gender, as well as their interactions with the MOOC platform. Among these interactions are the number of connections per day, the number of chapters read, the number of videos viewed, and interactions with the material resources available on the platform. There are several methods for collecting data, including:

**Facebook:** Through the Facebook API, data may be collected after authorization from its community by specifying certain criteria related to the exploration's purposes.

**Web scraping:** Web scraping consists of extracting data from any website using a program, automatic software, or another site in a structured format for reuse.

**Twitter:** With the Twint tool, we can collect every single detail regarding Twitter. The developer only needs to specify the desired information in the code.

**Preprocessing and feature selection.** The information collected can contain several types of errors, including typographical errors, missing information, inaccuracies, etc. Therefore, the preprocessing of data remains a crucial aspect of ensuring the collected information is accurate, consistent, and usable. In this case, it may be necessary to replace, modify, or delete data that contains incorrect information, significantly improving the data's reliability and consistency. We will also need to identify the most relevant features for our model. As a part of the feature selection phase, the number of input variables is reduced during the model development process. Generally, the best variables have a strong correlation with the target variable.

### 5.3 Profile identification

**Learner profile generation.** As a result of the preprocessing phase, learners will be grouped into profiles according to their skills. By using a clustering algorithm, such as K-means, learners in a profile are grouped according to similar characteristics to those in their group. Thus, the learning path can be customized by considering the choices of the learners surrounding the corresponding learner at risk.

**Learner at risk identification.** The purpose of this step is to detect learners who are at risk of dropping out of the MOOC by using classification algorithms. Based on the features previously selected, the algorithm will identify learners who are at risk of not completing the course. Using such an early warning model, we can provide learners with real-time risk assessments, ensuring that the appropriate support can be provided through a learning path that meets their skill levels and needs.

### 5.4 ACO algorithm principle

The choice of path to follow for each learner will be determined according to the ACO method inspired by the foraging behavior of ant colonies [21]. When an ant finds a certain amount of food, it deposits pheromones along the path of the nest based on the quantity and quality of the food. Therefore, other ants use the path marked by the trace, called the joint path leading to the food source. In this paper, we propose using the ants' ability to cooperate, communicate, and learn to determine the optimal path for each learner in the MOOC context. Formally, an ant will move from node *i* to node *j* with probability:

$$p_{ij}^{k} = \begin{cases} \dfrac{(\tau_{ij})^{\alpha} . (\eta_{ij})^{\beta}}{\sum_{l \in J_i^k} (\tau_{il})^{\alpha} . (\eta_{il})^{\beta}} & If\ j \in J_i^k \\ 0\ If\ j \notin J_i^k \end{cases} \tag{1}$$

$J_i^k$ is the set of neighbors of vertex i of the kth ant,
$\tau_{ij}$ is the amount of pheromone trail on edge (*i*, *j*),
$\alpha$ and $\beta$ are weightings that control the pheromone trail
$d_{ij}$ is the distance between vertices *i* and *j*.
The visibility value $\eta_{ij}$ is given by: $\eta_{ij} = \dfrac{1}{d_{ij}}$

The pheromone values are updated each iteration by all the m ants that have built a solution in the iteration itself. The pheromone $\tau_{ij}$, which is associated with the edge joining vertices *i* and *j*, is updated as follows:

$$(\tau_{ij}) = (1-\rho)\tau_{ij} + \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \quad (2)$$

ρ is the pheromone evaporation rate,
m is the number of ants,
$\Delta\tau_{ij}^{k}(t)$ is the quantity of pheromone laid on edgs ($i$, $j$) by the ant $k$:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L^k} & \text{If the ant } k \text{ used edge}(i,\ j) \text{ in its tours,} \\ 0 & \text{Otherwise} \end{cases} \quad (3)$$

Where: $Q$ is a constant and $L^k$ is the length of the tour constructed by the ant $k$.

## 6 ADAPTIVE LEARNING OPTIMIZATION: A CASE STUDY

Dropout rates in MOOCs can be prevented, contained, and even reversed if appropriate measures are taken. In this article, a solution is proposed for identifying at-risk learners and providing optimal learning paths tailored to their needs. In this case, learners with similar skills are grouped together to determine which courses to take using the efficiency of the ACO algorithm. The algorithm will then identify appropriate courses for each learner based on the experience of nearby learners.

### 6.1 OULAD dataset processing

**Description.** There has been a massive increase in the amount of student data in the online higher education sector. For instance, open university learning analytics (OULAD) contains a subset of open university (OU) student data collected in 2013 and 2014. The dataset contains information about 22 courses, 32,593 students, their assessment results, and logs of their interactions with the VLE summarized in daily summaries of student clicks (10,655,280 entries). The following is a summary of the OULAD data [22]:

**Code_module**: Module identification code for the student.
**Code_presentation**: Code of the presentation during which a learner is registered for the module.
**ID_student**: Learner identification number.
**Gender**: The gender of the learner.
**Region**: The location of the learner during the time of the module presentation.
**Highest_education**: The level of education that the learner possessed prior to the presentation.
**Age_band**: The learner age
Num_of_prev_attempts: This indicates how many times the learner has attempted to complete this module.
**Studied_credits**: The total number of credits for the modules the learner is currently enrolled in.
**Disability**: This indicates whether the learner has declared a disability.
**Final_result**: The learner's final results in the module presentation.

Furthermore, the number of clicks on virtual learning environment (VLE) activities that includes:

- **Resource**: Usually contains pdf resources such as books
- **Oucontent**: Represents content of assignments.
- **Forumng**: Discussion forum
- **Url**: Contains links to external/internal resources (video/audio content)
- **Homepage**: Course homepage
- **Subpage**: Points to other sites in the course together with basic instructions

**Data preprocessing.** Preprocessing improves the quality of data before it is processed. This is a preliminary step that brings together a set of preparatory phases that includes:

**Data cleaning:** Usually, the data contains noise and missing values; they are incomplete and sometimes in an unusable format. Cleaning is the process of removing these errors and restoring the data to a usable format.

**Encoding:** It converts alphanumeric data into numeric data. A variable is encoded using ordinal encoding by assigning a single decimal value to each category.

## 6.2 Features selection and normalization

The accuracy of models may be compromised if they include irrelevant features. Thus, selecting relevant features is one of the crucial steps in developing a specific predictor. Generally, the parameters used for feature selection are classified based on two factors: the similarity of the information and the quantum of information contributed by the features. For the selection of variables in our model, we use the similarity of information based on the correlation factor. In this case, we analyzed the relationship between our target (certification) and different variables (characteristics) using a distribution diagram separating certified and non-certified learners. The results are very interesting, showing a strong correlation between our target and the selected variables. In our context, we can show strong links between certain variables and our target by visualizing their relationship as illustrated in Figure 2.

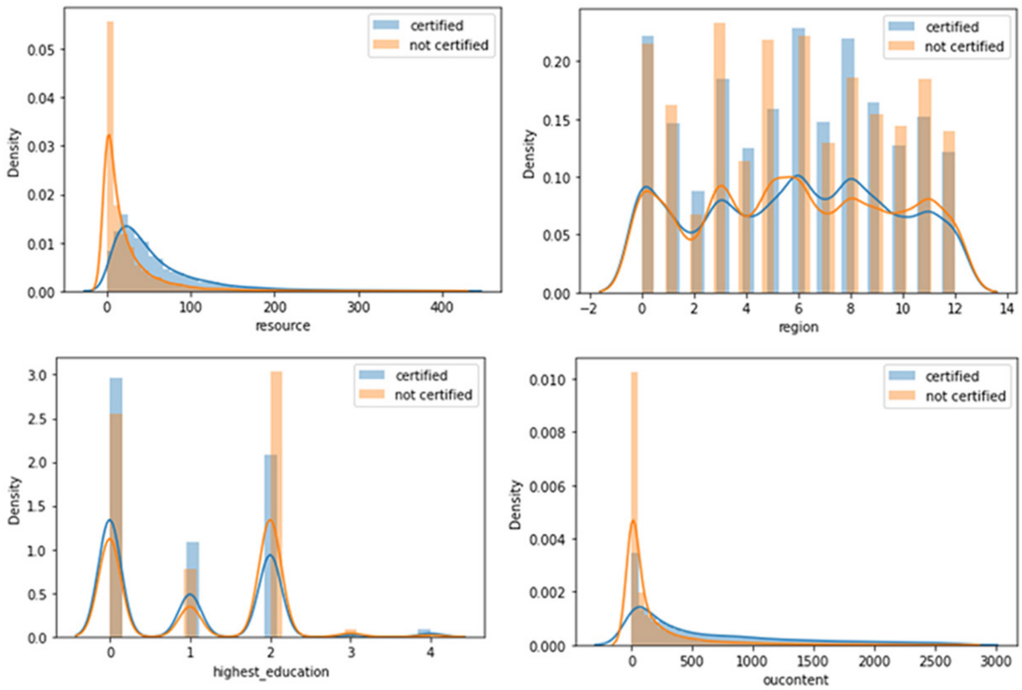



**Fig. 2.** *(Continued)*

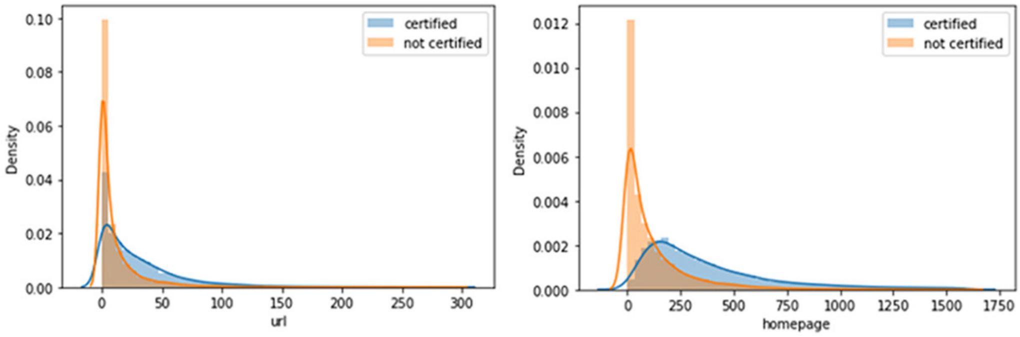



**Fig. 2.** Relationship between the certification and different features

According to the graphs above, there is a correlation between the variables resource, outcontent, url, homepage, subpage, forumng, dataplus, highest_education, region, and our target. In addition, it is essential to determine and quantify the extent to which the variables in our dataset are interdependent. Figure 3 represents the existing correlations between our variables (features).

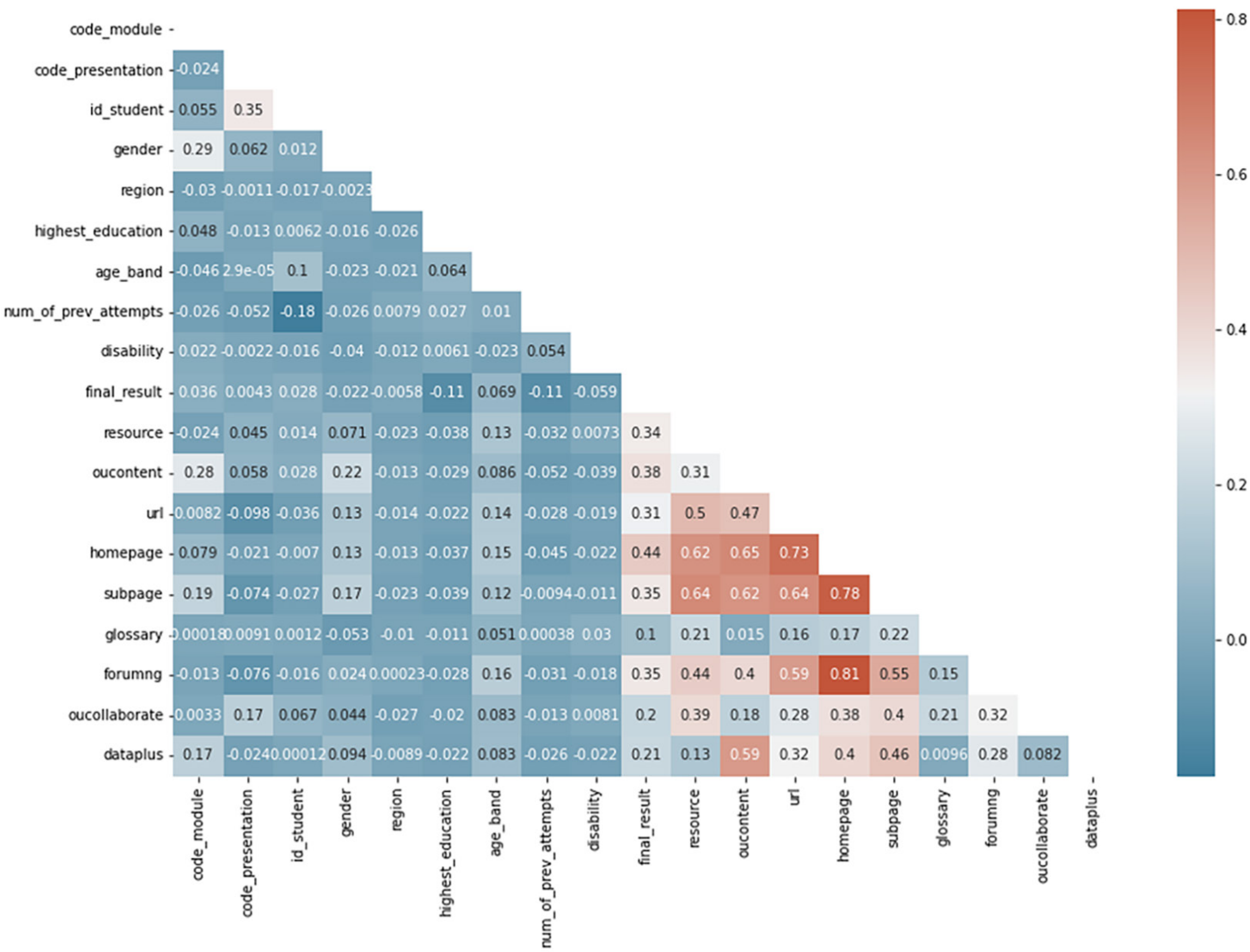



**Fig. 3.** Correlations between our features (features dependency)

Afterwards, we will proceed to eliminate other variables whose correlations are extremely close to zero, which are unlikely to be relevant to our model. This includes code_module, code_presentation, id_student, gender, region, highest_education, age_band, num_of_previous_attempts, and disability.

Therefore, the suggested model will include the following variables: homepage, content, subpage, resource, forum, url, and dataplus. Next, normalizing the data can significantly improve the performance of our model. Scaling is a common method for standardizing and involves transforming data so that each feature has a mean of 0 (zero) and a standard deviation of 1.

### 6.3 Prediction and optimal paths recommendation for at-risk learners

**Learners' at-risk prediction: profiles identification.** In order to group learners into profiles according to their skills, we suggest using the K-means algorithm. Furthermore, one of the most popular methods for determining the optimum number of clusters is the "elbow" method [23]. The method follows the evolution of the cost function of our model. The aim is to detect a zone of "bends" while minimizing the cost function. The elbow method runs multiple tests with different values for k (number of clusters) and records the score (variance measure) for each run. The Figure 4 below indicates the scores obtained for each k-value. A suitable solution minimizes the sum of the distances between cluster elements and their centroid (in our example, k = 2).

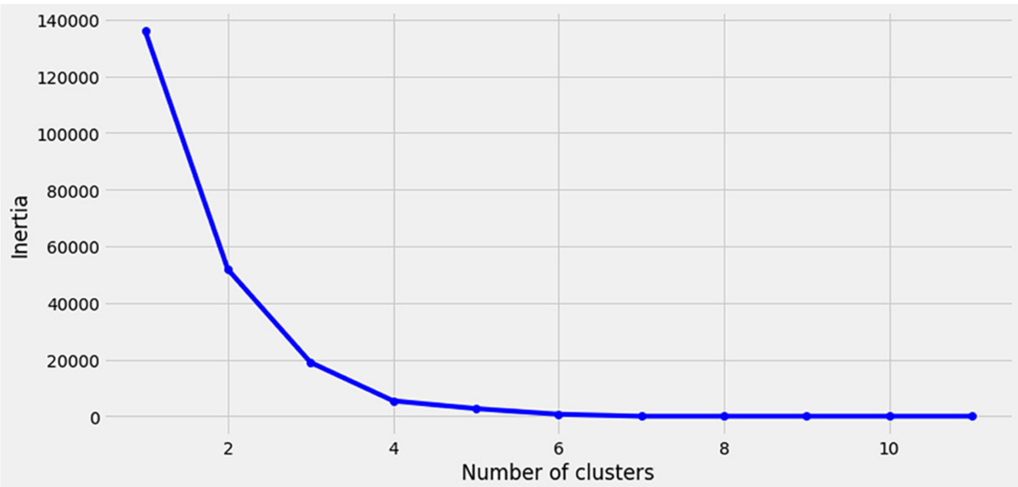


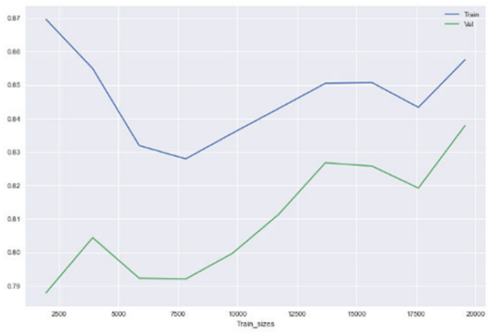

**Fig. 4.** Determination of the number of profiles using the elbow method

**Learners' at-risk prediction: classification algorithms evaluation.** As part of our effort to identify learners who are likely to abandon the MOOC, we subdivided our dataset into three sets of data: a Train_Set, a Test_Set, and a Validation_Set. Choosing an algorithm depends on which one returns the most accurate predictions (y_pred) as compared to expected values (y_test) and which yields the most accurate results. Initially, we train each model on the Train_Set. Then, we select the one that will have the most effective performance on the Validation_Set. Finally, we evaluate its performance on the Test_Set.

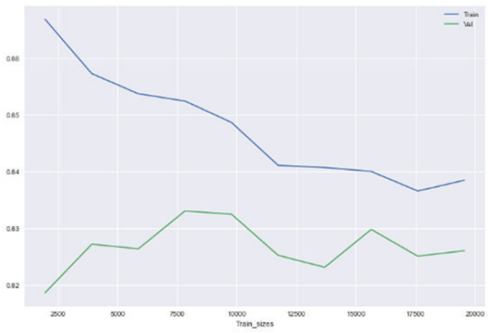
In order to evaluate the performance of each algorithm, we can use:

**Learning curves**: Pattern learning curves can be used during training to diagnose learning problems, such as pattern underfitting or overfitting, and to determine whether training and test data sets are sufficiently representative. As described in

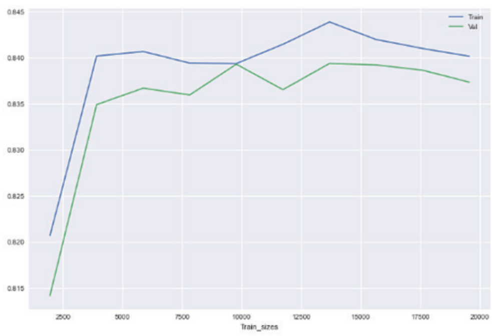
Figure 5a and b, we notice that the convergence of the learning scores and the validation scores requires a large amount of data when using the decision tree classifier algorithm and the XGBoost algorithm. However, for the AdaBoost and gradient boosting algorithms, convergence occurs as the size of ensemble learning increases, resulting in convergent learning scores and validation scores as illustrated in Figure 5c and d.



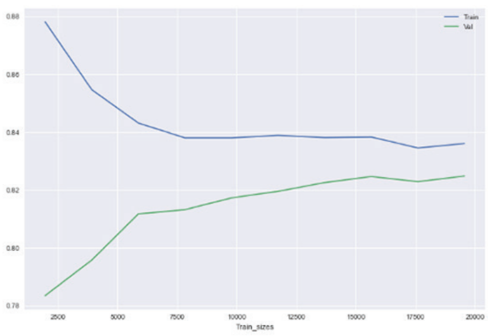

**(a) Decision Tree learning curve**




**(b) Extreme Gradient Boosting (XGBoost) learning curve**




**(c) Adaptive Boosting (AdaBoost) learning curve**




**(d) Gradient Boosting learning curve**

**Fig. 5.** Learning curves for various algorithms using OULAD dataset

Furthermore, the confusion matrix (Figure 6) shows that gradient boosting produces better results, as we correctly identified 2577 of the 3424 non-certified (at-risk) records, with 847 in error, which results in an overall recall of 75%. As for the adaptive boosting algorithm, only 970 cases at risk were correctly identified (recall 72%).

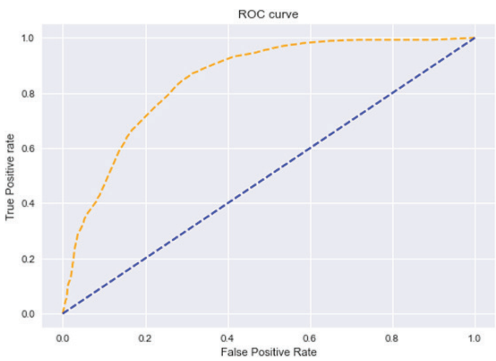
```
Adaptive Boosting (AdaBoost)

        [[2454  970]
         [ 520 2575]]
                      precision    recall  f1-score   support

                   0       0.83      0.72      0.77      3424
                   1       0.73      0.83      0.78      3095
```

```
Gradient Boosting

        [[2577  847]
         [ 556 2539]]
                      precision    recall  f1-score   support

                   0       0.82      0.75      0.79      3424
                   1       0.75      0.82      0.78      3095
```

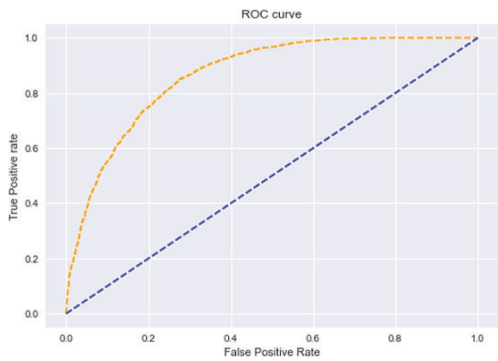
**Fig. 6.** Confusion matrix for AdaBoost and gradient boosting

**The AUC-ROC curve**: As part of our analysis, we used the AUC-ROC curve in order to visualize the performance of our model. In fact, ROC curves provide a measure of performance for classification problems at a variety of threshold parameters. Accordingly, ROC refers to a probability curve, and AUC indicates how well the model can distinguish between classes. In the ROC curve, true positive rate (TPR) is plotted on the Y axis, and false positive rate (FPR) is plotted on the X axis. Thus, the AUC-ROC curve summarizes the real trade-off between the real positive rate and the predictive value of a predictive model using different thresholds of probability, which is an incredibly essential aspect of classification problems.

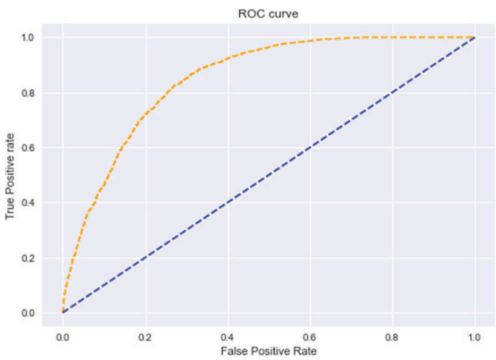
In our context, the higher the AUC, the better the model will be for distinguishing between "featured students" and "learners at risk." The scores calculated when using previous algorithms such as decision tree classifiers, XGBoost, AdaBoost, and gradient boosting, show that the gradient boosting algorithm is the most effective (86.63%), whereas the XGBoost algorithm obtained 86.56%, the AdaBoost algorithm (84.90%), and the decision tree algorithm (84.65%), as illustrated in Figure 7.



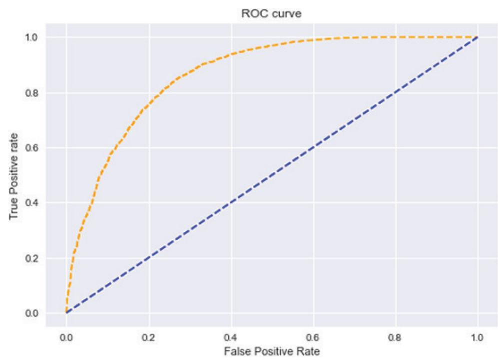

**ROC curve for Decision Tree**




**ROC curve for XGBoost**




**ROC curve for AdaBoost**




**ROC curve for Gradient Boosting**

**Fig. 7.** ROC curve for decision tree, XGBoost, AdaBoost, and gradient boosting

**Optimal paths recommendation for learners at-risk.** Optimizing the learner path in MOOCs is a major challenge for MOOC designers and teachers. The goal is to ensure that learners follow the most efficient path to achieve the course objectives and develop the desired skills. As part of our proposal, we suggest using adaptive learning to help personalize learner paths according to their particular abilities. As such, after identifying learners at risk of dropping out in the previous section, the learners are grouped according to their skills, and specialized courses are then offered.

For this purpose, we first use PCA (a powerful dimensionality reduction technique) [24] to transform input data into a smaller-dimensional space, i.e., a space with fewer variables. To reduce unnecessary complexity in our dataset. PCA is usually used for the visualization of multidimensional data as it reduces the number of dimensions while retaining important information. Figure 7 shows the projection of our two profiles in 2D space, with the centroid in red using dimensionality reduction (PCA).

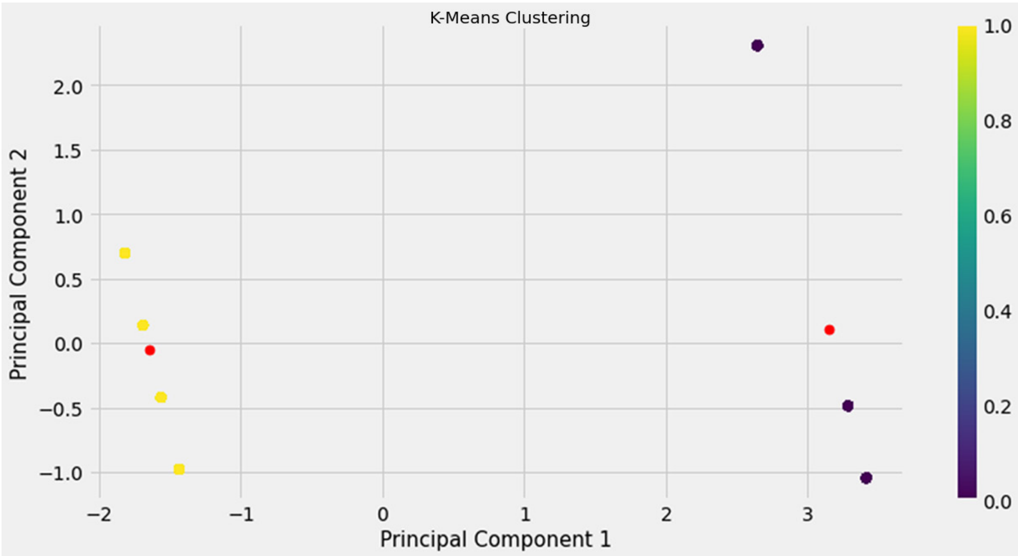



**Fig. 8.** The two profiles obtained using K-means algorithm

In this case, similar learners are grouped within the same cluster, and the distances between learners in the same cluster (intra-cluster) should be small as they have similar characteristics. Therefore, we can generate the learner's journey based on the courses taken by neighboring learners who share almost all of their characteristics.

Furthermore, ACO is useful to find an optimal path connecting courses within a cluster. The basic idea behind an ant-based algorithm is to use positive feedback mechanisms based on the laying of pheromones. The pheromone component keeps the most effective solutions found in memory, which can be used to develop better solutions.

In our case, the learners and their courses (represented by their coordinates (x, y)) would be the nodes, while the connections between the learners would be the edges, so courses previously viewed by the learner would not be offered. As such, we first identify the learner at risk and limit the search within the cluster (profile) to which the learner belongs. Then, the algorithm calculates the Euclidean distance between the learner and their neighboring learners after projection into two-dimensional space, which is the distance in a straight line between two nodes.

The algorithm returns a list of courses which represents the course to be suggested to the learner. For example, Figure 8 illustrates the use of the PCA method for clustering data presented in Table 1.

**Table 1.** Modules and learner's skills transformed into PCA components

| Component 1 | Component 2 | Module | Skills |
|---|---|---|---|
| –1.434007 | –0.981041 | CCC | Science, Technology, Engineering, and Mathematics |
| –1.562063 | –0.422329 | DDD | Science, Technology, Engineering, and Mathematics |
| –1.690120 | 0.136383 | EEE | Science, Technology, Engineering, and Mathematics |
| –1.818177 | 0.695095 | FFF | Science, Technology, Engineering, and Mathematics |
| 3.412289 | –1.046395 | AAA | Social Sciences |
| 3.284232 | –0.487683 | BBB | Social Sciences |
| 2.643949 | 2.305877 | GGG | Social Sciences |

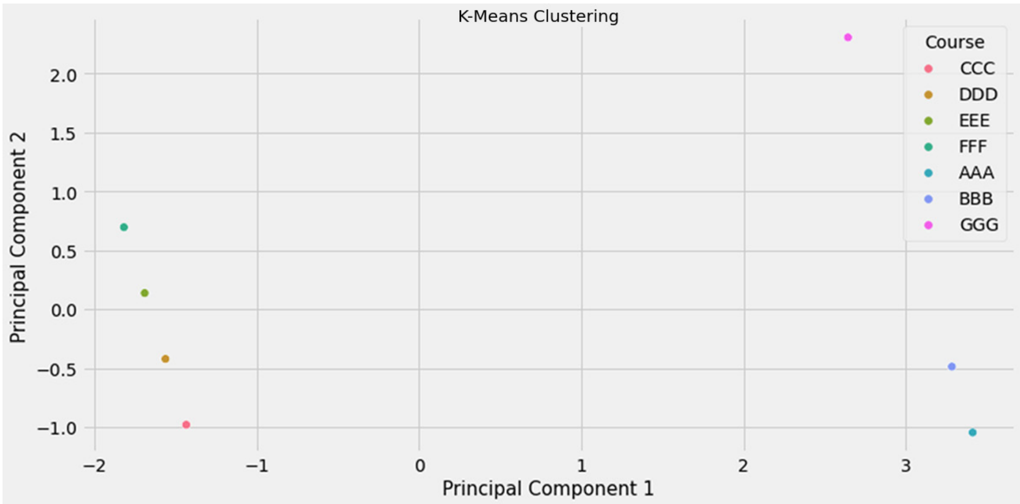



**Fig. 9.** The two profiles obtained using the K-means with the course labels

According to the distance between courses as shown in Figure 9, if the learner is interested in science, technology, engineering, and mathematics (STEM), the ACO algorithm will result in the following path: ['FFF', 'EEE', 'DDD', 'CCC', 'AAA', 'BBB', 'GGG'] whereas when we use the ACO algorithm for a learner who is interested in the field of social sciences, we will get the following path: ['GGG', 'BBB', 'AAAA', 'CCC', 'DDD', 'EEE', 'FFF'].

As such, the adaptation of courses in MOOCs will promote learner retention and improve their course completion rate. In fact, adapting paths in MOOCs can play a key role in retaining learners and improving their course completion rate. Our approach adapts the course in real time based on the experiences of other learners with similar profiles to meet the needs of at-risk learners. Course optimization and learner prediction, as described in this study, can help MOOC designers create more effective and personalized learning experiences. Additionally, machine learning techniques and associated algorithms can provide insight into learner behavior and allow course providers to optimize the learning experience for each learner.

Furthermore, several key factors must be considered when adapting MOOCs path in order to meet the needs and preferences of learners, in particular with regard to their level of knowledge and their objectives. It is very important to assess the level of knowledge and skills of the learners in the field of the course. This helps to better understand their strengths and shortcomings and to customize the courses accordingly. Learners may have different learning goals, such as acquiring specific skills, exploring a subject area, or preparing for certification. Some learners may require additional instructional support due to their level of knowledge or learning preferences. MOOC moderators should integrate complementary resources such as discussion forums, online tutorial sessions, or question-and-answer sessions to meet the specific needs of learners and guide them throughout their learning experience. In our case study, monitoring learner progress is still a key factor in providimg a pathway that meets these needs.

In addition, the adapted courses must be scalable and adjusted according to learner feedback and the changing needs of the target audience. MOOCs need to monitor completion rates, collect learner feedback, and conduct regular assessments to constantly improve the learning experience and ensure paths match learners' changing needs and preferences. By taking these factors into account, MOOCs can provide a more tailored and personalized learning experience, which contributes to learner satisfaction and the achievement of their learning goals.

## 7 CONCLUSION

In summary, using learners' interactions with the learning environment from the OULAD database, we propose two complementary solutions to reduce dropout rates in MOOCs. The first is to predict whether learners are at risk of dropping out of the MOOC program using an adaptive boosting algorithm. The second solution involves tailoring the learning experience for learners identified as at risk by creating profiles to support material selection and decision-making. Furthermore, course selection is conducted using the ACO optimization algorithm to find the path to follow based on learners' distance. As part of adaptive learning, our approach adapts the course in real-time based on the experiences of other learners with similar profiles to meet the needs of the at-risk learners. In this context, course optimization and learner prediction, as described in this study, can help MOOC designers create more effective and personalized learning experiences. Furthermore, machine learning techniques and associated algorithms can provide insight into learner behavior and enable course providers to optimize each learner's learning path.

As prospects, we intend to identify groups of users based on their learning styles through fuzzy clustering. Furthermore, we plan to evaluate the quality of recommendations generated by our system in light of feedback received from students.

## 8 REFERENCES

[1] M. Gaebel, "MOOCs: Massive open online courses," in *MOOCs: Massive Open Online Courses*, vol. 11, EUA Occasional Papers, 2013.

[2] P. Kerr, "Adaptive learning," *Elt. J.*, vol. 70, no. 1, pp. 88–93, 2016. https://doi.org/10.1093/elt/ccv055

[3] H. Chen, X. Wu, and J. Hu, "An improved K-means clustering algorithm," in *Proc. IEEE 3rd Int. Conf. Commun. Softw. Netw.*, pp. 44–46, 2011.

[4] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons B*, vol. 4, pp. 51–62, 2017. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05

[5] M. Gong, "A novel performance measure for machine learning classification," *Int. J. Manag. Inf. Technol.*, vol. 13, no. 1, pp. 1–19, 2021. https://doi.org/10.5121/ijmit.2021.13101

[6] Z. Kanetaki, C. Stergiou, G. Bekas, C. Troussas, and C. Sgouropoulou, "Analysis of engineering student data in online higher education during the COVID-19 pandemic," *International Journal of Engineering Pedagogy (iJEP)*, vol. 11, no. 6, pp. 27–49, 2021. https://doi.org/10.3991/ijep.v11i6.23259
[7] S. Jacques and T. Lequeu, "Remote knowledge acquisition and assessment during the COVID-19 pandemic," *Int. J. Eng. Pedagog.*, vol. 10, no. 6, pp. 120–138, 2020. https://doi.org/10.3991/ijep.v10i6.16205
[8] S. Jacques and T. Lequeu, "The attractiveness of reversing teaching forms—feedback on an electrical engineering course," *International Journal of Engineering Pedagogy (iJEP)*, vol. 10, no. 3, pp. 21–34, 2020. https://doi.org/10.3991/ijep.v10i3.12361
[9] Z. Kanetaki, C. Stergiou, G. Bekas, C. Troussas, and C. Sgouropoulou, "A hybrid machine learning model for grade prediction in online engineering education," *Int. J. Eng. Pedagogy (IJEP)*, vol. 12, no. 3, pp. 4–24, 2022. https://doi.org/10.3991/ijep.v12i3.23873
[10] M. Harrathi and R. Braham, "Recommenders in improving students' engagement in large scale open learning." *Procedia Computer Science*, vol. 192, pp. 1121–1131, 2021. https://doi.org/10.1016/j.procs.2021.08.115
[11] Q. Fu, Z. Gao, J. Zhou, and Y. Zheng, "CLSA: A novel deep learning model for MOOC dropout prediction," *Computers & Electrical Engineering*, vol. 94, p. 107315, 2021. https://doi.org/10.1016/j.compeleceng.2021.107315
[12] V. Vanitha, P. Krishnan, and R. Elakkiya, "Collaborative optimization algorithm for learning path construction in E-learning," *Computers and Electrical Engineering*, vol. 77, pp. 325–338, 2019. https://doi.org/10.1016/j.compeleceng.2019.06.016
[13] A. Ewais and D. A. Samra, "Adaptive MOOCs based on intended learning outcomes using naive bayesian technique," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 4, pp. 4–21, 2020. https://doi.org/10.3991/ijet.v15i04.11420
[14] R. Li, "Adaptive learning model based on ant colony algorithm", *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 1, pp. 49–57, 2019. https://doi.org/10.3991/ijet.v14i01.9487
[15] M. Kaouni, F. Lakrami, and O. Labouidya, "The design of an adaptive E-learning model based on artificial intelligence for enhancing online teaching," *International Journal of Emerging Technologies in Learning*, vol. 18, no. 6, p. 202, 2023. https://doi.org/10.3991/ijet.v18i06.35839
[16] S. Assami, N. Daoudi, and R. Ajhoun, "Implementation of a machine learning-based mooc recommender system using learner motivation prediction," *International Journal of Engineering Pedagogy*, vol. 12, no. 5, pp. 68–85, 2022. https://doi.org/10.3991/ijep.v12i5.30523
[17] E. M. Smaili, C. Khoudda, S. Sraidi, S. Azzouzi, and M. E. H. Charaf, "An innovative approach to prevent learners' dropout from MOOCs using optimal personalized learning paths: An online learning case study," *Statistics, Optimization & Information Computing*, vol. 10, no. 1, pp. 45–58, 2022. https://doi.org/10.19139/soic-2310-5070-1206
[18] S. Sraidi, E. M. Smaili, S. Azzouzi, and M. E. H. Charaf, "A neural network-based system to predict early MOOC dropout," *International Journal of Engineering Pedagogy (iJEP)*, vol. 12, no. 5, pp. 86–101, 2022. https://doi.org/10.3991/ijep.v12i5.33779
[19] E. M. Smaili, S. Sraidi, S. Azzouzi, and M. E. H. Charaf, "An optimized method for adaptive learning based on PSO Algorithm," in *IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, pp. 1–5, 2020. https://doi.org/10.1109/ICECOCS50124.2020.9314617
[20] S. Soukaina, S. El Miloud, and M. El Hassan Charaf, "MOOCs performance analysis based on quality and machine learning approaches," in *IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, Kenitra, Morocco, pp. 1–5, 2020. https://doi.org/10.1109/ICECOCS50124.2020.9314606

[21] J. Dréo, A. Pétrowski, P. Siarry, and E. Taillard, "Metaheuristics for hard optimization," Springer, 2006.

[22] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open university learning analytics dataset," *Sci. Data*, vol. 4, p. 170171, 2017. https://doi.org/10.1038/sdata.2017.171

[23] T. M. Kodinariya and P. R. Makwana, "Review on determining number of clusters in K-means clustering," *Int. J.*, vol. 1, no. 6, pp. 90–95, 2013.

[24] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9, pp. 2812–2831, 2014. https://doi.org/10.1039/C3AY41907J


## 9 AUTHORS


**El Miloud Smaili** is a PhD student in the field of machine learning/deep learning and holder of a state engineering degree in computer engineering at ENSAO, senior state engineer at the Faculty of Human and Social Sciences – University Ibn Tofail, Kenitra, Morocco (E-mail: miloud.smaili@uit.ac.ma).

**Mohamed Daoudi** is a Database administrator at Ibn Tofail University, Kenitra, Morocco since 2011, holder of a state engineer's degree in computer science at ENSAO and currently a PhD student in the field of mobile learning and learning analytics.

**Ilham Oumaira** is an Associate Professor at the National School of Applied Sciences-Kenitra. She is a member of the Laboratory of Engineering Science at Ibn Tofail University, Morocco. Since 2014, she has been in charge of the Center for Educational and Digital Innovation. Her current research focuses on MOOCs and more specifically on learning analytics.

**Salma Azzouzi** is a Professor at Ibn Tofail University – Faculty of Science (Computer Science Department). She is a member of the LaRI laboratory. She is also the general co-chair of the International Conference on Electronics, Control, Optimization and Computing (ICECOCS) and guest editor of a special issue of the journal SOIC. Her main areas of interest are: Distributed testing, e-learning, optimization (E-mail: salma.azzouzi@uit.ac.ma).

**Moulay El Hassan Charaf** is a Professor at Ibn Tofail University – Faculty of Sciences (Department of Computer Science). He is a member and joint director of the LaRI laboratory. He is also the general co-chair of the International Conference on Electronics, Control, Optimization and Computing (ICECOCS) and guest editor of the special issue on Intelligent Control for Future and Complex Systems in the International Journal of Modeling, Identification and Control (IJMIC). His main areas of interest are: Control, optimization, adaptive learning, distributed testing (E-mail: my.charaf@uit.ac.ma).