



Student Involvement in Mobile-Learning: Case of Ibn Tofail University

Mohamed DAOUDI^{1,*}, Nada LEBKIRI², Yassine Ouali¹, Ilham Oumaira¹

¹Laboratory of Electric, engineering, computing, and mathematical science laboratory, Ibn Tofail University, Morocco

²Telecommunications Systems and Decision Engineering Laboratory (LASTID), Ibn Tofail University, Morocco

Abstract In last years, the use of cell phones has reached new heights. This influences teaching methods at universities. The integration of mobile technologies into the teaching process can encourage the students to be more involved in the online learning process. The main challenges of mobile learning can be summarized in the changing attitudes in the educational field, being able to develop adequate pedagogical frameworks, good design (pedagogical and visual) and providing the right methods to control the involvement of the learners. Although mobile devices are highly present in the daily life of learners and trainers, the use of these technologies in distance education appears to still be low. The objective of this research is to measure the involvement of the students who are using smartphones compared to those who are using desktop computers by monitoring the learner's activity on the platform. To carry out this research, we used three Moodle distance-learning platforms from Ibn Tofail University to collect data. This data was processed by machine learning algorithms in an effort to see a link between the use of the mobile and the involvement of a student in an online learning.

Keywords Mobile Learning, Student involvement, Behavioral involvement, Machine Learning, Mobile Device, Moodle

AMS 2010 subject classifications 97U50, 97B40, 91C20

DOI: 10.19139/soic-2310-5070-1217

1. Introduction

In Moroccan universities, the great use of smartphones by the students to consult online courses has changed the way the courses 'content and presentation[1] are thought of, the effectiveness of learning differs depending on the type of the device as well as the involvement of the students in their studies.

Student involvement is the physical and psychological energy spent by the student in an academic learning experience [2]. On a distance-learning platform, behavioral involvement focuses on the learner's actions, such as number of connections, participation in forums, and completion of lessons and homework. Knowing that there are two types of involvement: behavioral and psychological involvement [3] and considering that we used the traces of the students on e-learning platforms to follow the behavior of the students, our study will only focus on behavioral involvement.

Most of the online courses are not designed and scripted to be received on mobile devices, which is a barrier that hinders the learning process [4]. M-learning facilities such as the internet, hardware, and software impact the readiness both directly and indirectly [5]. Despite these obstacles, mobile learning remains one of the main factors that encourages the involvement of students in online learning [6]. As a result, the following questions impose themselves:

- *How can Mobile Learning affect the Student's involvement?*

*Correspondence to: Mohamed DAOUDI (Email: mohamed.daoudi@uit.ac.ma). Laboratory of Electric, engineering, computing, and mathematical science laboratory, Ibn Tofail University, Kénitra, Morocco.

- *What are the factors that affect a learner's involvement in Mobile Learning?*

In this work, we study the student involvement in M-learning at Ibn Tofail University. Understanding the student's behavior at a distance-learning platform is an important step in studying the degree of learner involvement. We suggest the following hypotheses:

Hypothesis 1: Using Mobile devices in learning is positively related to Behavioral student involvement.

Hypothesis 2: Students use their smartphones in learning are more involved than students using desktops.

The structure of this research is as follows: section 2 and 3 presents a complete background and summary of the literature review relating to student involvement in mobile learning and research methodology are presented in the section 4. Section 5 presents the analysis of data collection and pre-processing. The result of our research and the discussion are addressed in section 6. Finally, the conclusion and perspectives of this work are presented in section 7.

2. Related work

In [7], the authors explored the effect of using a mobile phone on students' learning during a class lecture. The individuals in three separate study groups (control, low distraction, and high distraction) watched a video lecture, took notes on the lecture, and completed two learning assessments after watching the lecture. The results of this study showed that students who did not use their mobile phones noted 62% more information in their notes, were able to recall more detailed information from the lecture, and scored 1.5 letters higher on a multiple-choice test than students who actively used their cell phones.

A research framework has been developed and has been empirically tested on the basis of on data collected from 309 college students using mobile learning on an online education platform in China [8]. The results showed that cognitive and affective involvement in learning has positive and significant effects on students' pursuit of mobile learning. The implications for practitioners are simple: mobile learning providers and educators should consider learners' cognitive and affective involvement in promoting their mobile learning platform.

The authors of 'Student engagement in mobile learning via text message' [9] examined how 93,819 Kenyan students use a text message-based mobile learning platform with millions of users in sub-Saharan Africa. The study focused on longitudinal variation in engagement over a one-year period for students in different age groups and tested for evidence of learning gains using learning curve analysis. The results of this study showed that the student's engagement is highest during school vacations and near standardized exams, but persistence over time is low (less than 25% of students return to the platform after enrolling). On the other hand, by grouping students into three groups based on their activity level, the study examined the variation in students' learning behaviors and test scores over the first ten days. Highly active students show promising trends in terms of quiz completion.

In the same direction and in order to evaluate students' involvement, critical thinking and attitudes towards collaborative learning, the study of [10], evaluated students' learning in three different collaborative learning environments, with and without mobile technology. The results indicate that mobile technology is associated with positive student perceptions of collaborative learning, but increased student dis-involvement during class. In addition, students' level of critical thinking was more closely associated with the tools used to construct written responses than with the style of collaborative learning environment. Students who constructed responses in paragraph form on a mobile device demonstrated significantly lower critical thinking than those who used a computer keyboard or wrote their responses by hand.

The study of [11] focuses on the notion of student re-involvement in the context of a mobile learning platform, how it is predicted, and how it differs from dis-involvement. The authors analyzed data from 1,196,780 quiz attempts from 87,651 Kenyan students, and find that 36.3% of students who disengage for a week or more eventually re-engage on the platform. They spend more time on the quizzes initially than students who remain disengaged. A Random Forest classifier trained on two days of student activity predicts disinvolvement and re-involvement with similar performance: F1 scores are 81.2% and 80.9%, respectively.

Several research has focused on understanding the behaviors related to student involvement in MOOCs [12] [13] [14] [15] [16]. The authors [18] were able to meaningfully classify student types and visualize patterns of student involvement that were not previously clear. The results of this research contribute to the educational community's understanding of student involvement and performance in MOOCs, and provide the broader learning analytics community with suggestions for new ways to approach the analysis and visualization of learning analytics data.

The related work mentioned above focuses on student involvement in mobile learning environments. The results show that mobile learning can build a barrier that hinders the learning process and the results are not of the same quality as traditional online learning via desktop computers with a large screen and keyboard. We propose this work because most of it does not detail the factors that negatively or positively affect involvement in this type of learning. For example, do the technical characteristics of the mobile device such as screen size, device memory contribute to student disinvolvement.

The results of this research contribute, on the one hand, to the educational community understanding the relationship between student involvement and the use of mobile devices in the Moodle platform. On the other hand, trying to remove the parameters that influence the student's behavioral involvement in Mobile learning.

3. Preliminaries

In this research, we have exploited data from the Moodle platform, which is exploited with a machine learning k-means algorithm in order to measure the involvement of students who use mobile devices. Before starting our research, a presentation of the key concepts seems important to us.

3.1. Moodle platform

The word Moodle is originally an acronym for Modular Object-Oriented Dynamic Learning Environment, which is especially useful to programmers and educational theorists. Moodle is a powerful, open-source learning management system that allows users to create robust, flexible and engaging online learning experiences. It is an alternative to proprietary commercial e-learning solutions [13]. Moodle was created by Martin Dougiamas an educator in computer science who spent time supporting a CMS at a university in Australia [14]. Moodle has been successfully installed in institutions and universities around the world [13]. A learning organization has full control over the source code and can make modifications as remquired. The modular design of Moodle makes it very easy to create new courses, by adding content that will engage learners, and it is designed to support a learning style called social constructivist pedagogy [[15] [16].

3.2. Mobile learning

Mobile learning is defined as "any educational delivery whose unique or dominant technologies are handheld or portable devices." This definition means that mobile learning could include cell phones, smartphones, personal digital assistants (PDAs) and their peripherals, perhaps tablet PCs, and perhaps laptop PCs, but not desktop computers in carts and similar solutions. Perhaps the definition should also take into account the growing number of experiments with dedicated mobile devices, such as game consoles and iPods, and encompass both mainstream industrial technologies and ad hoc experimental technologies [17] [22] . With the uncertainty of whether laptops and tablets enable mobile learning, Figure 1 illustrates the difficulty of this definition.

In this work, we mean by Desktop PC or Laptop and by Mobile Smartphone or Tablet. These are the devices most used by students to consult Moodle platforms.

3.3. Student involvement

Student involvement refers to the amount of physical and psychological energy that the student devotes to the academic experience [18].

According to Astin (1999), an example of a highly involved student is one who:

- Devotes considerable energy to studying;

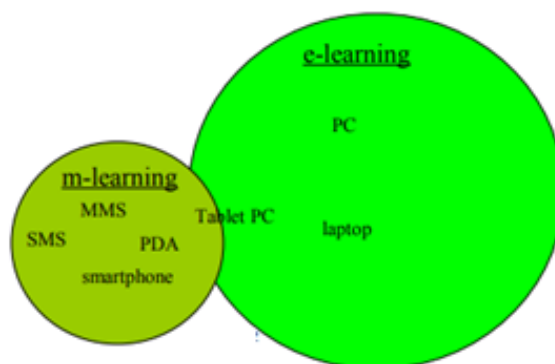


Figure 1. M-learning vs. E-learning [18]

- Spends much time on campus;
- Participates actively in student organizations;
- Interacts frequently with faculty members and other students;

4. Research Methodology

To carry out this research, the chosen methodology was to provide a response to the various elements of our problematic and to fulfill our research objectives. The current study proposed that the use of mobiles in E-learning would positively affect the affective involvement of learners, which will allow a continuous mobile learning experience.

4.1. Educational Choices

Moodle is an online learning platform. In order to allow a load balance, Ibn Tofail University has set up three Moodle platforms. Each platform is intended for a population that belongs to the same field of study. The use of these three platforms will allow us to compare the results of three populations from different fields.

In such a learning environment, the system must be able to allow us to define the learner profile and the learning context. The extraction of traces and data generated by the three platforms Moodle will be deployed with a learning analytics approach based on the platforms databases. Several performance indicators will be analyzed such as the type of mobile device, connection time, activity, Etc. These extracted variables make it possible to measure the student's involvement in the platform.

4.2. Technical Choices

A Machine Learning K-means algorithm will then analyze the data collected from the three Moodle platforms. This will allow us to identify learner's cluster with similar characteristics and to investigate the relationship between the use of a device and the involvement of the learners. Below, we will present the operation of the K-means algorithm and the notion of clustering.

Data clustering

An operational definition of clustering can be stated as follows: Given a representation of n objects, find K groups based on a measure of similarity such that the similarities between objects in the same group are high while the similarities between objects in different groups are low. The main clustering objectives are:

Table 1.

algorithms performance.

Platform one dataset			Platform two dataset			Platform three dataset			The average of the three datasets		
Silhouette Coefficient	Calinski-Harabasz Index	Davies-Bouldin Index	Silhouette Coefficient	Calinski-Harabasz Index	Davies-Bouldin Index	Silhouette Coefficient	Calinski-Harabasz Index	Davies-Bouldin Index	Silhouette Coefficient	Calinski-Harabasz Index	Davies-Bouldin Index
0,2873	868,0558	1,3662	0,3876	2373,4178	1,1828	0,4386	1119,5824	0,9587	0,371166667	1453,685333	1,169233333
0,2606	760,7758	1,4345	0,3452	2047,4781	1,3265	0,4028	1030,3379	1,2239	0,3362	1279,5306	1,3283
0,3711	61,4579	0,9647	0,1563	143,1396	1,1139	0,4684	196,4849	1,0082	0,331933333	133,6941333	1,028933333
0,2468	425,9918	1,5695	0,3831	1880,5601	1,4061	0,49	1232,561	1,0993	0,3733	1179,7043	1,3583

- Underlying structure: grouping of data, detection of anomalies and making hypotheses to understand the data structure.
- Natural classification: to identify similarities in the data to deduce phylogenetic relationships. Compression: grouping data and presenting it in a more organized way to create prototypes.

Algorithm choice justification

Among the most popular algorithms for unsupervised clustering, there is K-means, Agglomerative clustering, Density-based spatial clustering (DBSCAN) and Gaussian Mixture Modeling (GMM). To choose the best clustering algorithm for our dataset, we must choose an algorithm that allows a best cluster compactness (points in the same cluster should be similar) and the separation (points in different clusters should be dissimilar). We will use the following measures:

- **The silhouette score** of a clustering is in $[-1, 1]$ and should be maximized [20].
- **The Davies-Bouldin score** of a clustering is in $[0, +\infty]$ and should be minimized [20].
- **The Calinski- Harabasz score** of a clustering is in $[0, +\infty]$ and should be maximized [20].

In table 1, we calculated the three scores of the three validation measures above for the three datasets that we will use during this research. Next, we calculated the average of the scores for the three datasets. The result shows that K-means is the best algorithm for our case. He had the best score for Silhouette Coefficient (0.371166667) and Calinski-Harabasz Index (1453.685333) and for Davies-Bouldin Index he had almost the same score with DBSCAN algorithm (1.169233333 vs 1.028933333).

K-means Algorithm

Clustering is an unsupervised learning method that will find patterns in the data. In particular, by grouping things that look-alike. Machine learning algorithms that use this learning method does not try to learn a correlation relationship between a set of features X of observation and a value to predict Y, as is the case for learning supervised. One of the best known clustering algorithms is k-means [21]. It is an analytic technique identifies groups of objects based on the proximity of objects in the center of k groups (map). The center is determined as the arithmetic mean of the n-dimensional attribute vector of each cluster (reduce). The main limitation of this method is the choice of an optimal number of clusters.

In unsupervised learning, the data is represented as follows:

$$X = \begin{pmatrix} x_{(1,1)} & x_{(1,2)} & x_{(1,...)} & x_{(1,n)} \\ x_{(2,1)} & x_{(2,2)} & x_{(2,...)} & x_{(2,n)} \\ \dots & \dots & \dots & \dots \\ x_{(m,1)} & x_{(m,2)} & x_{(m,...)} & x_{(m,n)} \end{pmatrix}$$

Below, the structure of the k-means algorithm:

Input

X: The dataset

Set K : The desired number of groups

Output

partition of K groups $\{C_1, C_2, \dots, C_k\}$

Start

1) A random initialization of the C_k centers;

Repeat

2) Affection: generate a new partition by affecting each object to the closest center group;

$$x_i \in C_k \text{ si } \forall_j |x_i - \mu_k| = \min_j |x_i - \mu_j| \quad \mu_k \text{ the center of the } K\text{-class};$$

3) Representation: Calculate the centers assigned to the new partition;

$$\mu_k = \frac{1}{N} \sum_{x_i \in C_k} x_i \quad \text{Until the algorithm converges to a stable partition};$$

End.

The main idea is to randomly choose a set of fixed centers and search iteratively for the optimal partition. Each individual is assigned to the nearest center, after all data has been assigned, the average for each group is calculated, it constitutes the new representatives of the groups, when a stationary state has been reached (no data changes group) the algorithm is stopped.

4.3. Users Experience in Mobile Learning Environment

In order to define the characteristics of a learner, we will try to model the interactions between different contexts and the profile of the learner in a mobile learning experience. This will allow us to choose the right parameters in the data extraction part. Figure 2 shows the interactions between different contexts and the learner profile in a mobile learning experience [18].

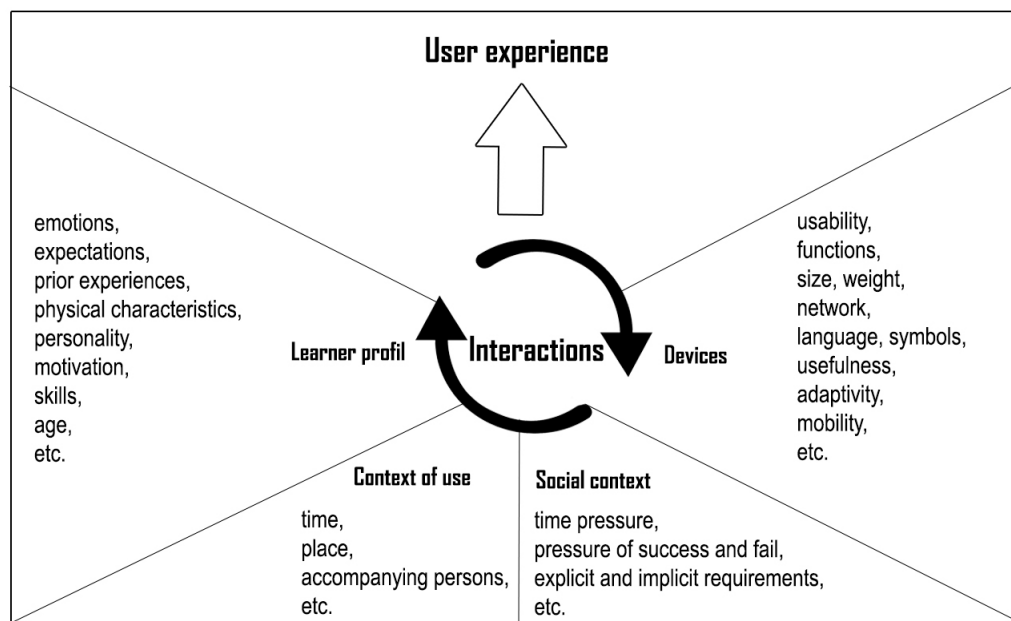


Figure 2. Users experience in mobile learning environment.

4.4. Workflow of our machine learning approach

Representation and clustering are the main steps in our approach, where we will represent and classify learners using k-means algorithm based on Learner's profile structure in mobile learning and the context profile presented in the previous section. After the representation and the classification of learners, the second stage of the approach

is keeping the parameters that are strongly correlated with student's involvement and remove the other parameters [18].

The evaluation of the relevance of the parameters used in our model will be the objective of the third step of our approach [18].

The last step of our approach is the optimization step. At this level, an improvement of our k-means algorithm is necessary to correct the problems observed during the analysis of the learner's traces at the evaluation stage. The next step is to propose new model with new parameters and more data [18].

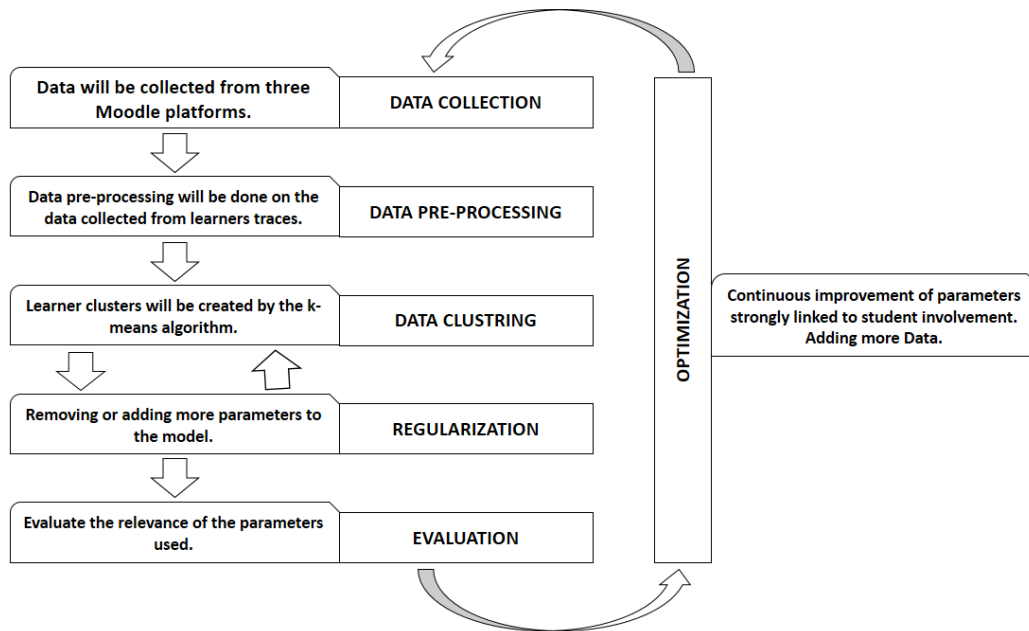


Figure 3. Workflow of our machine learning approach.

5. Data collection & data preprocessing

5.1. Data Collection

We started our research by collecting data from Moodle platform of the National School of Applied Sciences. The same processing is carried out on the three databases. To extract the data, SQL query was developed.

In the Moodle database, the system logs all actions of a user in log tables. 'logstore_standard_log' record information such as actions taken by a user for each course, quiz, chat, etc. The 'user_device' table is used to store information on the mobile devices used by the students to connect to the platform. The 'user' table, which contains information of the learner. Figure 4 shows the UML class diagram representing data structure used in our research. This diagram is created based on the MERISE data model published on the official Moodle website.

To understand the behavior of the learner in online learning scenario, we will use the analytic tables of the Moodle platform database. The learners' traces on the Moodle databases can be the mobile devices type, registration information, periods of connections, the activities of the learner, etc. Figure 5 shows the information stored of devices' types used by the learner to connect to the Moodle platform and the standard log table, which records all student activities on the platform. This dataset is at the level of one row per student, per course. So, for example, if a student signed up for three courses, that person would have three lines associated with their ID on

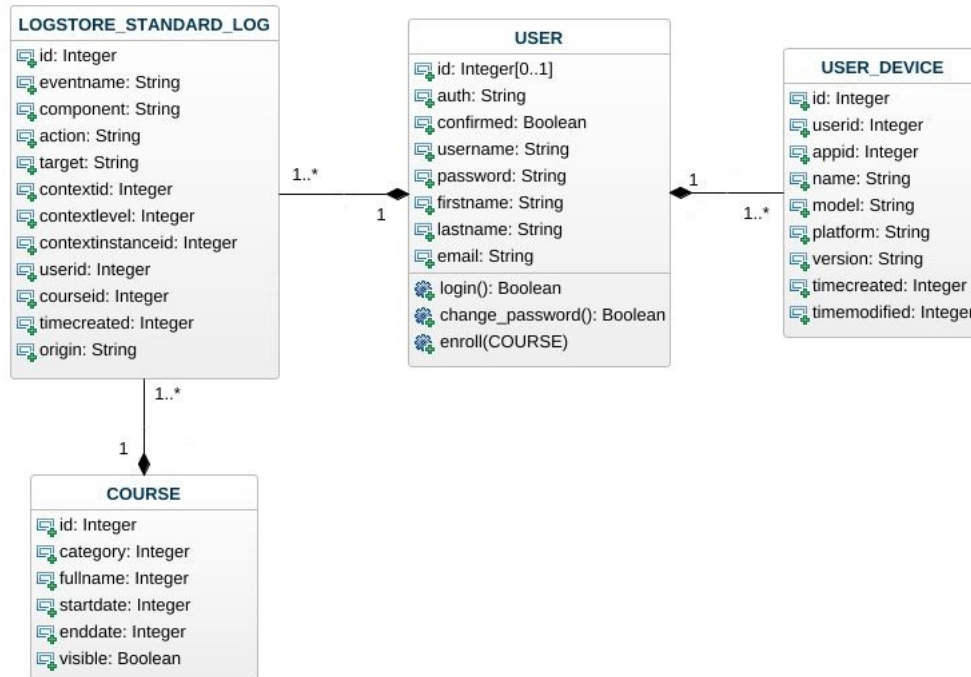


Figure 4. UML class diagram representing data structure.

the database. For a detailed description of the variables included in the de-identified data set, please see the list of variables below.

id_ligne	Interface	Origin	Device	Device_type	user_id	Course_ID	Score	Pass	file_size_MB	action	nbr_action	Activity	nbr_send_msg	nbr_cha
1	Smartphone	web	Apple	iPhone8,1	14870	276.0	100.0	Yes	1569813.0	viewed	1280000	1280000	0	
2	PC	web	DESKTOP	DESKTOP	14662	NaN	100.0	Yes	484198.0	viewed	1267200	1267200	0	
3	PC	web	DESKTOP	DESKTOP	14662	276.0	100.0	Yes	484198.0	viewed	1267200	1267200	0	
4	PC	web	DESKTOP	DESKTOP	14662	137.0	100.0	Yes	484198.0	viewed	1267200	1267200	0	
5	Smartphone	Mobile	Apple	iPhone8,1	14870	276.0	100.0	Yes	1434907.0	viewed	1170000	1170000	0	

Figure 5. learner Dataset.

Variable definitions:

- **Interface:** the interface offered by the system (Mobile / desktop);
- **Origin:** Connection source, 'ws' from the mobile application or 'web' from a web browser;
- **Device:** The connection device (Desktop, Samsung, Apple, etc.);
- **Device_type:** The model of the device.
- **user_id:** The Student ID;
- **Course_ID:** The Course ID;
- **Pass:** State of progress of the student in a course;
- **file_size_MB:** Size of a user's files;
- **action:** the type of action performed by the user (viewed, created, downloaded, etc.);
- **nbr_action:** number of actions performed by the user;

Table 2.

Size of collected data.			
	Faculty of Sciences platform	National School of Applied Sciences platform	Faculty of Languages, Letters and Arts platform
Dataset size	2583	3328	6854

- **Activity:** total activities of a user;
- **nbr_send_msg:** number of messages sent to the platform;
- **nbr_chat_msg:** number of chat messages sent to the platform;
- **Parts_of_the_Day:** The part of days when user activity is performed;
- **attempt_state:** attempt state in a quiz;
- **attempt_grade:** Final state of progress in a quiz;

The data collection allowed us to have significant data size. The table below shows the dataset size for each platform.

5.2. Data Preprocessing

Data preprocessing is a crucial step in preparing raw data. Typically, native data contains outliers, null values, or in an inappropriate format that cannot be used in machine learning models. The cleaning of this data is necessary. For the data preprocessing, we followed the following steps:

Importing libraries

In order to perform data preprocessing using Python, it would be necessary to import predefined Python libraries.

- **Numpy:** is used to include large multidimensional arrays and matrices.
- **Matplotlib:** is a Python 2D plotting library.
- **Pandas:** is used for importing and managing the datasets.
- **Seaborn:** is used for data visualization based on matplotlib.
- **Scikit-learn:** provides unsupervised and supervised learning algorithms.

Importing datasets

The data collected from the three Moodle platforms are exported in CSV format. The dataset will be imported with Pandas for data preparation. Without prior manual processing, data is imported as is and exported from Moodle databases.

Finding Missing Data

At this step, for each feature of our dataset, we have to determine the number of missing values. Figure 6 shows the number of missing values for each feature.

We noticed that six variables in our dataset contain missing values. Since the number of these values is high, we decided to remove these variables from our dataset and keep the variable *'file_size_MB'* which we considered important to measure the activity of the learner on the platform. To have a fully populated dataset, we decided to fill the null values of the *'file_size_MB'* variable with median values.

Encoding Categorical Data

The objective of this step is to transform non-numeric values into numeric ones. For this, we used the python's Scikit-Learn library to encode our dataset. Figure 7 shows an extract of our data after encoding. In our dataset, there are six categorical variables, interface, origin, device, Device_type, Pass, and Parts_of_the_Day. They are objects, which are not represented by numerical values.

Before				After	
user_final_grade	3319			Parts_of_the_Day	0
attempt_grade	3319			Activity	0
attempt_state	3283			nbr_action	0
Course_ID	2008			action	0
Score	1657			file_size_MB	0
file_size_MB	1651			Pass	0
Activity	0			Score	0
nbr_action	0			Course_ID	0
action	0			Device_type	0
nbr_send_msg	0			Device	0
Pass	0			Origin	0
nbr_chat_msg	0			Interface	0
Parts_of_the_Day	0				
user_id	0				
Device_type	0				
Device	0				
Origin	0				
Interface	0				

Figure 6. Number of missing values for each feature, after and before dropping.

id_ligne	Interface	Origin	Device	Device_type	Course_ID	Score	Pass	file_size_MB	action	nbr_action	Activity	Parts_of_the_Day
1	1	1	0	32	11	4	1	1569813.0	19	1280000	1280000	1
2	0	1	1	6	11	4	1	484198.0	19	1267200	1267200	1
3	0	1	1	6	11	4	1	484198.0	19	1267200	1267200	1
4	0	1	1	6	4	4	1	484198.0	19	1267200	1267200	1
5	1	0	0	32	11	4	1	1434907.0	19	1170000	1170000	2

Figure 7. Dataset after encoding.

Feature scaling

Scaling the characteristics of our dataset is the last step in data preprocessing. This technique allows the standardization of independent variables of the dataset in an interval. La figure 8 show our data after encoding.

id_ligne	Interface	Origin	Device	Device_type	Course_ID	Score	Pass	file_size_MB	action	nbr_action	Activity	Parts_of_the_Day
1	0.061649	0.005857	-0.026540	0.106007	0.006140	0.001319	0.027557	0.630314	0.037567	0.540020	0.541358	0.005263
2	-0.015834	0.007483	-0.010865	-0.011435	0.007846	0.001685	0.035211	0.245586	0.048001	0.683067	0.684762	0.006725
3	-0.015834	0.007483	-0.010865	-0.011435	0.007846	0.001685	0.035211	0.245586	0.048001	0.683067	0.684762	0.006725
4	-0.015766	0.007451	-0.010819	-0.011386	-0.092684	0.001678	0.035061	0.244536	0.047796	0.680147	0.681835	0.006697
5	0.066632	-0.140988	-0.028685	0.114575	0.006637	0.001426	0.029784	0.622419	0.040603	0.533169	0.534502	0.042238
...
3324	0.359927	-0.761571	0.476872	0.180073	0.035850	0.007701	-0.161856	-0.018257	0.030228	-0.021496	-0.020803	0.030730
3325	0.343425	-0.726654	0.455008	0.171816	0.285644	0.007348	-0.154435	-0.017420	-0.100034	-0.020510	-0.019849	0.029321
3326	0.358127	-0.757763	0.474487	0.179172	0.035671	0.007663	-0.161047	-0.018166	-0.104317	-0.021388	-0.020699	0.030576
3327	0.527489	0.050111	0.698877	0.263904	0.052540	0.011286	-0.237207	-0.026757	-0.312009	-0.031503	-0.030488	0.045035
3328	0.483603	0.045942	0.640731	0.241948	0.402237	0.010347	-0.217472	-0.024531	-0.286050	-0.028882	-0.027952	0.041289

Figure 8. Dataset after standardization.

Detection and elimination of outliers

Outliers in a dataset results in poor fit and poorer predictive modeling performance. Before passing our data to the machine learning program, we must remove the outliers from our dataset. The figure 9 shows the presence of outliers for the three variables *file_size_MB*, *nbr_action* and *activity*. We used the value of z-score to remove outliers from our dataset.

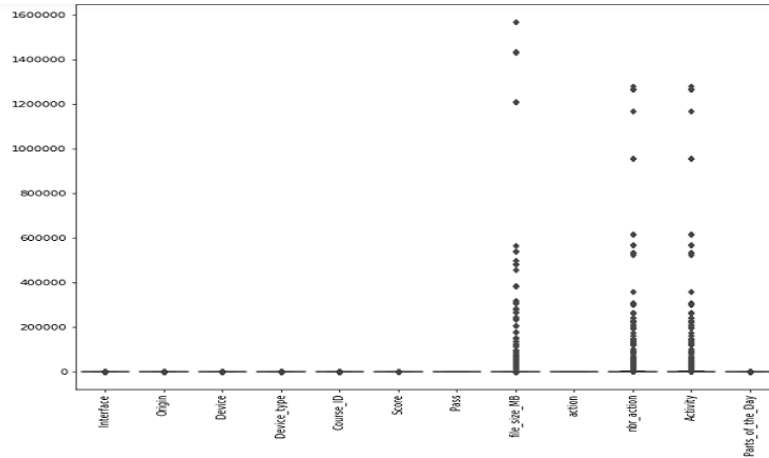


Figure 9. Dataset outliers.

We used the value of z-score to remove outliers from our dataset. In statistics, a z-score tells us how many standard deviations away a value is from the mean. We use the following formula to calculate a z-score[23].

$$z = (X - \mu) / \sigma$$

6. Results & Discussion

Correlation matrix

Since our dataset contains several columns, to check the correlations between the columns, we visualized the correlation matrix as a heatmap (figure 10).

Result: What drew our attention to the correlation matrix is the existence of a weak but significant correlation between the device used by the learner and his activity on the platform. More particularly with the feature *file_size_MB*, *nbr_action* and *activity*.

Discussion: This signifies the existence of a relationship between the involvement of the student represented by his activities on the platform and the device he uses to connect. Adding some feature related to the devices used as screen size and memory can help explain this relationship.

Choosing the Optimal Number of Clusters

The number of clusters (k) is the most important hyperparameter in K-Means clustering[24]. there are several methods to find the optimal value of k. we used two of these methods:

- Elbow Method multidimensional arrays and matrices.
- Silhouette Method

Our database has several dimensions, which makes it very difficult to visualize. As a result, the optimal number of clusters is no longer obvious. Fortunately, we have a way to determine this mathematically.

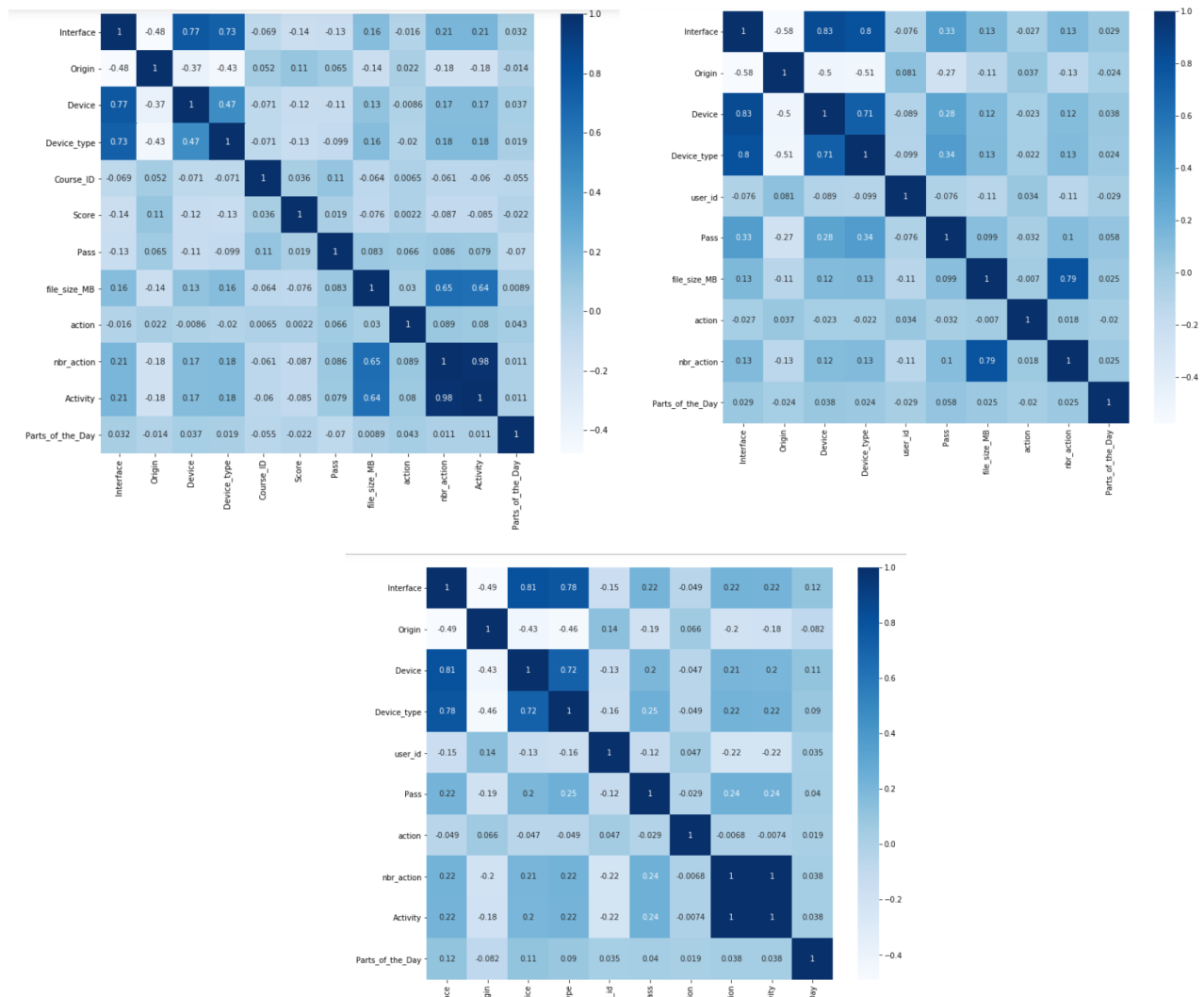


Figure 10. Correlation matrix of the three datasets.

Elbow and Silhouette methods are used to find the optimal number of clusters. For the Silhouette method and for a cluster number equal to 6, all the plots are more or less of similar thickness and hence are of similar sizes, as can be considered as best 'k'. For the Elbow method, we graph the relationship between the number of clusters and the sum of squares within clusters (WCSS), and then we select the number of clusters where the change in WCSS begins to stabilize (six clusters).

Result: we will categorize the data using the optimal number of clusters (6) that we are determined in the last step.

Discussion: For the three platforms: the Elbow method allows you to have six clusters. This means that the features of our dataset allow the population to be distributed over six categories. The next step in this research is to export each row with the corresponding cluster and do a deep analysis to find out what are the similar characteristics of each cluster.

K-means clustering

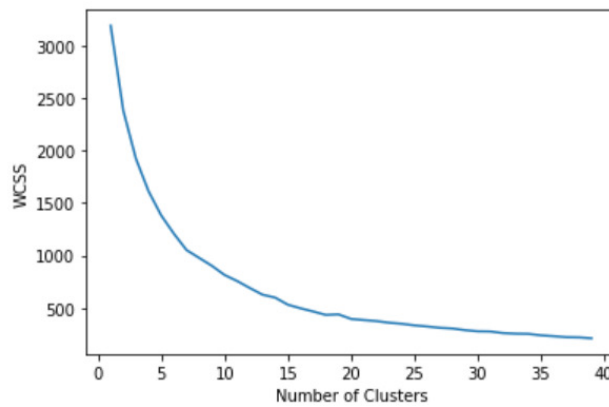


Figure 11. Elbow method graph.

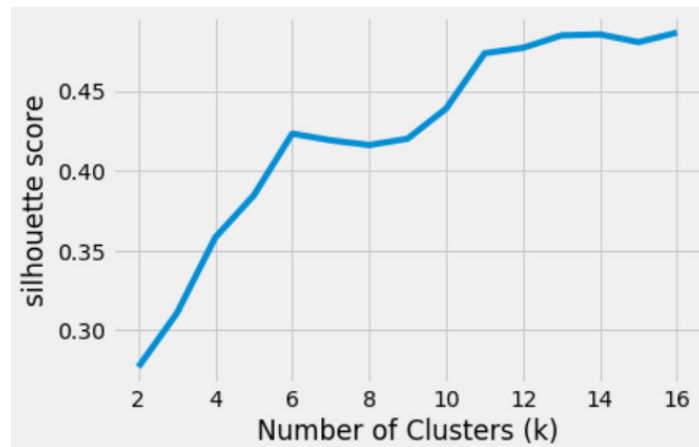


Figure 12. Silhouette Method graph.

Result: The execution of the k-means algorithm displays the result in the figure 12. Six different categories are messed up. This was not possible without using a clustering algorithm such as k-means.

Reading the content of each cluster shows homogeneity. For example, the majority of students who use mobile devices are grouped together in a single cluster for the three platforms.

Discussion: We noticed that, for the three platforms, the Elbow method allows having six clusters. This means that the population we are studying is a bit heterogeneous. In the next step of this research, we will work on a single course and with a homogeneous population (same class) to be able to focus only on the involvement of students who use a mobile device in online learning.

K-means clustering

Result: Our analysis of data extracted from the three Moodle platforms is focused on understanding student behavioral involvement. For our study, student behavior is measured by the number of actions (loggedin, viewed, sent, created, downloaded, etc.) performed by the student within the platform. A student who is more involved than another is a student who has performed more actions. The result of this analysis is shown in Table 3.

Discussion: Except for the platform of the National School of Applied Sciences, the result of the table above shows that students who use a PC / Laptop perform more actions on the platform than those who use smartphones or tablets. This partially refutes our second hypothesis. This finding may prompt us to look for other factors that influence student behavioral involvement in Mobile learning. This result may be due to the mastery of the use of

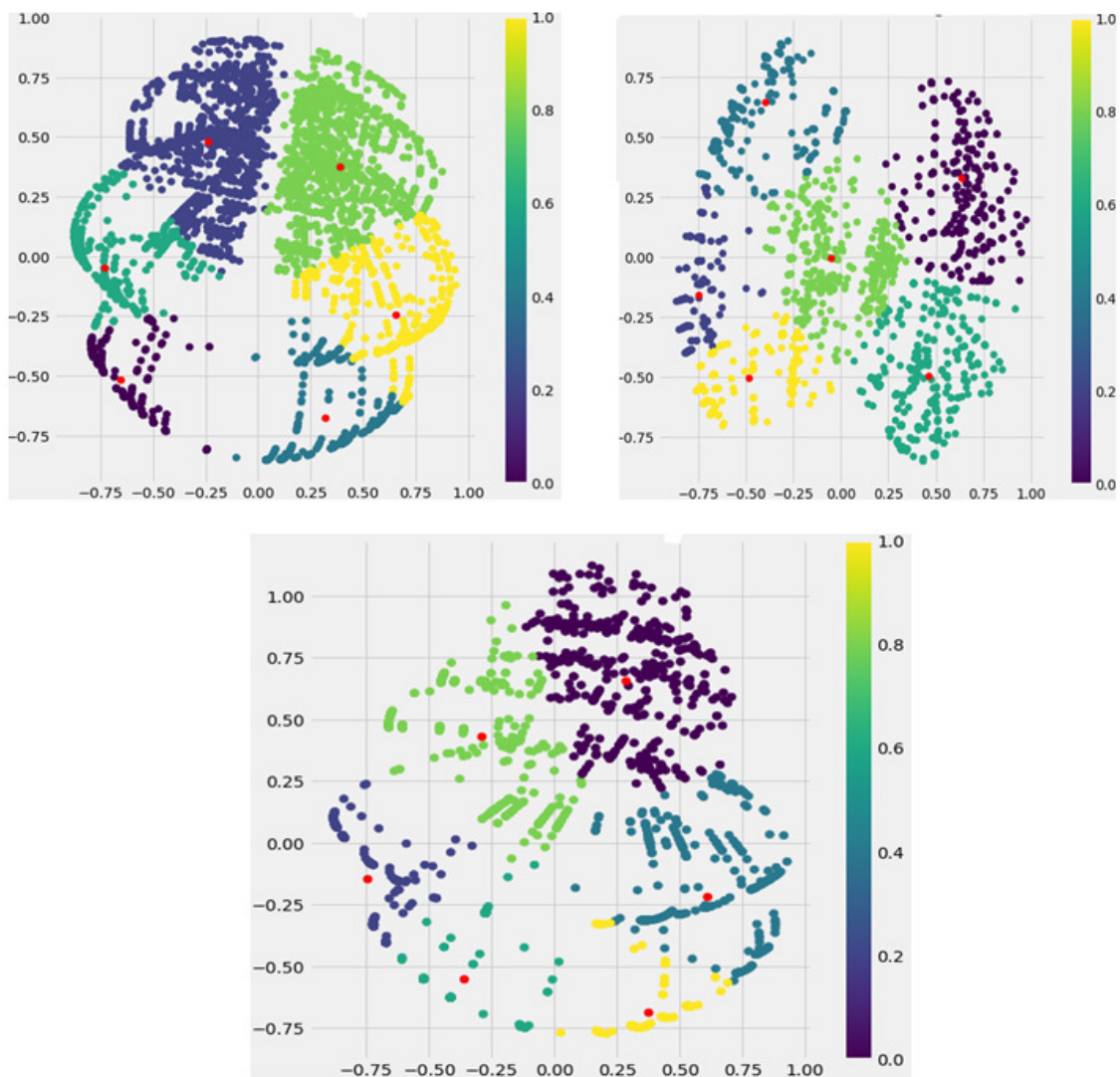


Figure 13. K-means clusters of three datasets.

Table 3.

Student behavior on Moodle platforms.

	Faculty of Languages, Letters and Arts Moodle platform		Faculty of Sciences Moodle platform		National School of Applied Sciences Moodle platform	
	Average of actions taken by a student	Average size of uploaded files (Ko)	Average of actions taken by a student	Average size of uploaded files (Ko)	Average of actions taken by a student	Average size of uploaded files (Ko)
Desktop	1078,09955	433,353194	1867,4021	1619,48581	120,27332	Not calculated
Smartphone	793,3619247	267,5648536	12678,5171	4061,80541	170,593482	Not calculated

smartphones by students, because the students of the National School of Applied Sciences are always confronted with computer media such as Smartphones and Tablets which the opposite for the students of Faculty of Languages, Letters and Arts.

By analyzing the clusters of our k-means algorithm, we noticed that, for the three platforms, the students who use mobile devices are classified in an independent cluster with a few students who use PC / Laptop. This result prompted us to study the student's behavioral involvement by comparing this involvement with the mobile support used by the student. Among the constraints that restrain the use of smartphones are the size of the screen and the size of the RAM memory. We decided to add, for each entry of the cluster containing the students who use smartphones, two other parameters: the screen size and the size of the RAM memory of the smartphones.

Involvement & Mobile characteristics

Result: A first reading, after adding the two parameters: screen size and RAM size to our cluster, which contains students who use smartphones, shows that students with less efficient mobile devices performed fewer actions by comparing them to those who have a more powerful smartphone.

Discussion: This is a result that we have not confirmed and constitutes a new research track.

7. Conclusion

In this paper, we presented a K-means prediction model to analyze the involvement of students at Ibn Tofail University in learning via mobile according to data extraction of three MOODLE platforms: Moodle platform of the Faculty of Languages, Letters and Arts, Moodle platform of the Faculty of Sciences and Moodle platform of the National School of Applied Sciences. The results of this research showed that there is a clear link between the use of Mobile / Desktop devices and the involvement of learners in an online course.

The future work is now oriented towards an experiment that we have started, it requires focusing the study on the population of the same field in order to target the variables that directly affect the use of the mobile and the involvement of the student.

REFERENCES

1. Y. Zidoun, F. El Arroum, M. Talea, and R. Dehbi, *Students' perception about mobile learning in Morocco: Survey analysis*, Int. J. Interact. Mob. Technol., vol. 10, no. 4, pp. 80–84, 2016.
2. A. Astin, *Student Involvement: A Development Theory for Higher Education*, J. Coll. Stud. Dev., vol. 40, pp. 518–529, Jan. 1984.
3. D. S. Carlson and M. R. Frone, *Relation of behavioral and psychological involvement to a new four-factor conceptualization of work-family interference*, J. Bus. Psychol., vol. 17, no. 4, pp. 515–535, 2003.
4. S. J. Baldwin and Y.-H. Ching, *Guidelines for Designing Online Courses for Mobile Devices*, TechTrends, vol. 64, no. 3, pp. 413–422, 2020.
5. M. Z. Asghar, E. Barberà, I. Younas, *Mobile learning technology readiness and acceptance among pre-service teachers in Pakistan during the COVID-19 pandemic*, Knowl. Manag. E-Learn. Int. J., vol. 13, no. 1, Art. no. 1, Apr. 2021.
6. M. Ibrahim, N. Khairudin, and D. Salleh, *Unwinding Environment in the Lecture Room Using Mobile Learning (M-Learning) Adoption among the Millennial Generation*, vol. 1874, no. 1, 2021.
7. J. H. Kuznekoff and S. Titsworth, *The Impact of Mobile Phone Usage on Student Learning*, Commun. Educ., vol. 62, no. 3, pp. 233–252, Jul. 2013.
8. S. Yang, S. Zhou, and X. Cheng, *Why do college students continue to use mobile learning? Learning involvement and self-determination theory: College students mobile learning continuance*, Br. J. Educ. Technol., vol. 50, no. 2, pp. 626–637, Mar. 2019.
9. R. F. Kizilcec and M. Chen, *Student engagement in mobile learning via text message*, Proceedings of the Seventh ACM Conference on Learning@ Scale, pp. 157–166, 2020.
10. H. Heflin, J. Shewmaker, and J. Nguyen, *Impact of mobile technology on student attitudes, engagement, and learning*, Comput. Educ., vol. 107, pp. 91–99, Apr. 2017.
11. M. Chen and R. F. Kizilcec, *Return of the Student: Predicting Re-Engagement in Mobile Learning*, EDM, 2020.
12. C. Milligan, A. Littlejohn, and A. Margaryan, *Patterns of engagement in connectivist MOOCs*, J. Online Learn. Teach., vol. 9, no. 2, pp. 149–159, 2013.
13. R. Ferguson and D. Clow, *Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs)*, in Proceedings of the fifth international conference on learning analytics and knowledge, pp. 51–58, 2015.
14. S. I. De Freitas, J. Morgan, and D. Gibson, *Will MOOCs transform learning and teaching in higher education? Engagement and course retention in online learning provision*, Br. J. Educ. Technol., vol. 46, no. 3, pp. 455–471, 2015.

15. C. Taylor, K. Veeramachaneni, and U.-M. O'Reilly, *Likely to stop? predicting stopout in massive open online courses*, ArXiv Prepr. ArXiv14083382, 2014.
16. R. Longadge and S. Dongre, *Class imbalance problem in data mining review*, ArXiv Prepr. ArXiv13051707, 2013.
17. C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy, *Visualizing patterns of student engagement and performance in MOOCs*, Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14, Indianapolis, Indiana, pp. 83–92, 2014.
18. J. Cole and H. Foster, *Using Moodle: Teaching with the popular open source course management system.*, O'Reilly Media, Inc., 2007.
19. W. Rice, *Moodle 1.9 E-Learning Course Development*, Packt Publishing Ltd, 2008.
20. Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, *Understanding of Internal Clustering Validation Measures*, 2010 IEEE International Conference on Data Mining, pp. 911–916, Dec. 2010.
21. M. Z. Rodriguez et al., *Clustering algorithms: A comparative approach*, PLOS ONE, vol. 14, no. 1, p. e0210236, Jan. 2019.
22. Rice, William and William, H, *Moodle*, Packt publishing Birmingham, 2006.
23. zach, *How to Calculate Z-Scores in Python*, Statology, Jul. 03, 2020. Available from <https://www.statology.org/z-score-python> accessed Mar. 31, 2021).
24. S. Kumar, *Silhouette Method — Better than Elbow Method to find Optimal Clusters*, Medium, Aug. 09, 2021. Available from <https://towardsdatascience.com/silhouette-method-better-than-elbow-method-to-find-optimal-clusters-378d62ff6891> (accessed Sep. 19, 2021).
25. J. Traxler, *Defining mobile learning,* in *IADIS International Conference Mobile Learning*, pp. 261–266, 2005.
26. M. Sharples, I. Arnedillo-Sánchez, M. Milrad, and G. Vavoula, *Mobile Learning*, Technology-Enhanced Learning, pp. 233–249, 2009.
27. C. Romero, S. Ventura, and E. García, *Data mining in course management systems: Moodle case study and tutorial*, Comput. Educ., vol. 51, no. 1, pp. 368–384, Aug. 2008.
28. M. DAOUDI, N. LEBKIRI, and I. OUMAIRA, *Determining the Learner's Profile and Context Profile in Order to Propose Adaptive Mobile Interfaces Based on Machine Learning*, pp. 1–6, Dec. 2020.