Received September 4, 2018, accepted September 29, 2018, date of publication October 22, 2018,
date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2876755

Analyzing Learners Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach

RAGHAD AL-SHABANDAR[1], ABIR JAAFAR HUSSAIN[1], (Member, IEEE), PANOS LIATSIS[2], (Senior Member, IEEE), AND ROBERT KEIGHT[1]
[1]Faculty of Engineering and Technology, Liverpool John Moores University, Liverpool L33AF, U.K.
[2]Department of Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates
Corresponding author: Abir Jaafar Hussain (a.hussain@ljmu.ac.uk)

ABSTRACT Massive open online courses (MOOCs) have been experiencing increasing use and popularity in highly ranked universities in recent years. The opportunity of accessing high quality courseware content within such platforms, while eliminating the burden of educational, financial, and geographical obstacles has led to a rapid growth in participant numbers. The increasing number and diversity of participating learners has opened up new horizons to the research community for the investigation of effective learning environments. Learning Analytics has been used to investigate the impact of engagement on student performance. However, the extensive literature review indicates that there is little research on the impact of MOOCs, particularly in analyzing the link between behavioral engagement and motivation as predictors of learning outcomes. In this paper, we consider a dataset, which originates from online courses provided by Harvard University and the Massachusetts Institute of Technology, delivered through the edX platform. Two sets of empirical experiments are conducted using both statistical and machine learning techniques. Statistical methods are used to examine the association between engagement level and performance, including the consideration of learner educational backgrounds. The results indicate a significant gap between success and failure outcome learner groups, where successful learners are found to read and watch course material to a higher degree. Machine learning algorithms are used to automatically detect learners who are lacking in motivation at an early time in the course, thus providing instructors with insight in regards to student withdrawal.

INDEX TERMS Machine learning, massive open online courses, statistical analysis, big data.

I. INTRODUCTION

Online education is becoming increasingly widespread within the higher education context. There were more than 6 million students enrolled in online courses in 2012 [2]. The new bellwether of online educational platforms is Massive Open Online Courses (MOOCs) [3]. MOOCs are open educational platforms that deliver learning resources through digital platforms [4]. The reduction and potential of elimination financial, geographical, and educational obstacles led to a growing number of learners undertaking online courses. As of late 2012, global universities are offering a number of academic courses through commercial platforms such as HarvardX, Khan Academy, Coursera, and Udacity [2].

A variety of resources are used in such courses, including video lectures, weekly quizzes, regular assessments, and even PDF documents. Additionally, a learner can interact asynchronously with the instructors via postings in discussion forums. The increased number of enrolled users in MOOCs provides an opportunity for researchers to understand and analyze learner interactions with the online learning environment [2].

Learning Analytics (LA) has been used to gain deeper insight into course curriculums, course structure design, in addition to learner success and failure [5]. LA itself is an efficient analytics tool used by researchers to enhance and develop learning strategies. One of its distinctive features is the ability to analyze log data from online courses in an advanced fashion [5].

LA has been utilized to investigate the reasons behind participant enrollment in online classes through the analysis of student engagement patterns [6]. The findings demonstrate that students engage in online courses for two main

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

reasons, namely, feeling immediate satisfaction when undertaking a task, and attaining formal recognition by obtaining a certificate. Additionally, students have the possibility for flexible engagement in high-quality course settings without additional financial overhead [7]. A notable limitation of existing studies is the consideration of engagement style to student performance, without accounting for the level of student engagement as a factor in influencing learner participation [8].

Motivation has a significant impact on the development of the students' cognitive skills and in enhancing their performance. As such, highly motivated students are goal-oriented individuals, who tend to expand their experience and overcome challenges [9]. In the online context, research indicated that most online learners are intrinsically motivated rather than extrinsically motivated [10]. Although motivation plays an important role in the online learning context, a limited number of contemporary studies considered behavioural activity interplay factors that could affect participant motivation [11].

In this research, we examine the links between engagement, performance, and motivation, in the context of geographical influences. We employ LA tools in evaluating the links between the learners' educational background, engagement level and performance. Moreover, machine learning models are applied in the prediction of learner motivational status. Hence, the predictors in our experiments are based on quantitative log data, rather than questionnaire responses. Until now, most studies neglected the use of machine learning tools for analyzing the effect of learner motivation on engagement.

The aim of the experiments in this study is to analyze and evaluate log data that reflect learner activity. This analysis will facilitate instructors in designing future online courses to enhance student participation. In addition, the findings will provide educators with insight into the association of learning style and academic achievement. Finally, the experiments will provide an indicative case study to highlight the value of learning analytics and machine learning tools in the educational context.

II. LITERATURE REVIEW

A. ENGAGEMENT IN ONLINE LEARNING ENVIRONMENTS

Student engagement is considered an important prerequisite for learning in the online context, impact on performance, motivation, and attrition [12]. Engagement can be classified into three main categories, namely, behavioural, emotional, and cognitive engagement. Emotional engagement occurs when students feel emotionally engaged in a learning activity. Cognitive engagement refers to the students' feelings in regards to progress in the academic task, while behavioural engagement refers to the level of student participation in the learning activity [12].

Behavioural engagement is concerned with student behavioural activities. The absence of behavioural engagement could negatively influence student academic outcomes [12]. Behavioural engagement is considered a crucial factor in increasing concentration, persistence, and social interaction, ultimately resulting in improvements in student performance.

Learner engagement has been widely investigated in online learning. Coffrin *et al.* [3] employed learning analytics techniques in analysing the patterns of participant engagement in MOOCs. The number of video hits and assignment submissions was used as features in the assessment of completion rates. The results showed that only 29% of participants completed their assignments, whereas more than 60% viewed the associated videos [3].

Videos and assessments were used to describe the prototypical patterns of learners' engagement in the Coursera platform on a weekly basis. Four patterns of engagement were introduced, namely completing, auditing, disengagement, and sampling [13]. The k-means clustering algorithm was used to find subpopulations in the engagement patterns, with results indicating that most learners engage with the course for the purpose of watching video lectures [13]. Classifying students based on students' interaction with videos is suitable for any MOOCs platform that considers only videos lectures and assessments. Consequently, the narrow focus on the use of these features imposes limitations on the generality of the proposed approaches [14]. Other researchers examined the factors relevant to the structural aspects of MOOCs design that could raise the level of participant engagement [15]. Learner comments were used to validate how instructional design promotes student engagement. The authors' findings indicated that course material, interaction, and persistent monitoring of participant progress are critical elements in increasing the level of engagement [15].

Balakrishnan and Coetzee [16] employed Hidden Markov Models (HMM) in predicting student persistence in online courses. Courses were split into six-time intervals considering multiple behavioural features such as the number of videos viewed and the number of post threads on the course forum. The results revealed that approximately 1% of the students who watched at least 50% of videos dropped from the course [16]. In addition, the results indicated that students who do not participate in the course forum are more likely to withdraw from the course. Hence, the authors demonstrated that HMM provides deep insights into issues affecting student retention rates.

Probabilistic Soft Logic (PSL) was proposed in [17] to model student engagement. PSL is defined as a paradigm for developing probabilistic models. PSL uses first-order logic rules to represent variables in the model. Three types of engagement were defined in this study, namely, active and passive engagement and disengagement. The learners' activity was defined as active, when learners demonstrated interaction with the course such as posting on the discussion forums and submitting assignments. The label of passive engagement was assigned to learners who accessed the resources homepage, without proceeding further to specific forms of interaction, such as voting on a post, watching lectures and following discussion forums. Disengaged learners were defined as those who tended to quit from an online course. The authors of the study also examined the links between learner engagement and performance. The findings indicated that latent engagement enhances the performance of predictive models. As such, the PSL model, which accommodated for latent engagement, achieved higher performance than the model without latent engagement. The value of the Area Under the Curve (AUC) metric was equal to 0.7492 for the PSL model with latent variables, while the AUC acquired a value of 0.7393 for the PSL model without latent variables [17]. The study also found that inferring latent variables could help instructors understand the reasons behind poor student performance.

B. INCENTIVE MOTIVATION THEORY

Incentive Motivation Theory (IM) is a behaviourist theory of motivation developed by Skinner [18]. IM seeks to explain why human activity occurs relative to goals. IM theory introduces the notion of ''ramifications'', which are posited to be the basis for task-focused incentives. In particular, ramifications are classified into the main subtypes of tangible and intangible. Motivation categories are further explained in terms of three main dimensions, i.e., intrinsic incentive motivation, extrinsic incentive motivation, and amotivation [19]. Intrinsic motivation is attained from a student's perception of a task as interesting, challenging, and enjoyable. In contrast, extrinsic motivation originates from the expectation of rewards that lie outside of the activity itself .Intrinsically motivated students feel immediate satisfaction while undertaking a task. Conversely, extrinsically motivated students derive satisfaction from extrinsic reward mechanisms, such as attaining favorable exam marks or social rewards. Amotivation is another category of motivation, where the lack of incentives represents a key factor in student dropout [18], [19].

C. MOTIVATION IN ONLINE COURSES

In terms of education, motivation is described as a conceptual construct that directs and improves student behaviour towards a specific goal [10].

Current studies highlighted the importance of motivation as a factor in learner engagement. Much of the research reported in the literature focuses on the validation of motivational indicators within the setting of online courses. Osborne and Jones [20] found a strong correlation between motivation and domain identification within MOOCs, e.g., job prospects, knowledge expansion, social development, etc. The authors demonstrated that social factors play an important role in increasing student engagement and enhancing cognitive skills.

To validate motivation in MOOCs, several studies have designed questionnaire frameworks based on the Glynn scale, e.g., ''Science Motivation Questionnaire II''. In [21], the authors employed the Glynn scale to evaluate four types of motivation, namely, intrinsic motivation, self-determination, self-efficacy, and career motivation, comparing English with Arabic participants within the Coursera platform [9]. The results revealed a similar pattern of motivation categories for both English and Arabic participants within the studied setting [9]. The Situational Motivational Scale (SIM) was adopted in [11] to measure learner motivation in two teacher education courses, delivered by the New Zealand Tertiary Institution. Four subtypes of motivations were assessed in these studies, namely intrinsic motivation, external regulation, identified regulation, and amotivation. The students were asked to respond to 16 SIM questions related to particular assignments. The results demonstrated that participants in both case studies exhibited high levels of identified regulation and intrinsic motivation [11].

Other studies investigated how motivation can positively influence learner performance. For example, de Barba *et al.* [22] demonstrated that motivation has a significant impact on learner participation. Learning Analytics was used to evaluate learner participation and performance in Coursera. The authors utilized video hits and quiz attempts as features, serving as an indicator of learner participation. The results showed that the most successful participants tend to be intrinsically motivated [22]. In another study, sentiment analysis of participants' interview transcripts within the Coursera platform was adopted to investigate the association between motivation and engagement [23]. Acquired knowledge and work were reported as the main factors of influence for learner motivation in online course participation. In this work, learner experience was found to be a critical factor affecting engagement and motivation levels. Learners with higher levels of education were more likely to engage than those with less formal education, as they were found to have the ability to overcome barriers including technical and subject difficulty [23].

According to Cho and Heron [10], Self-Regulated Learning (SRL) is a key factor for the achievement of motivation in learning. The SRL framework identifies student control, autonomy in the learning process, and time management as factors for successful goal achievement. A highly autonomous approach towards learning is a distinctive characteristic of self-regulated learners. Cho and Heron examined SLR in the context of motivation and learning strategy in an online mathematics course. The results indicated that learning delivery strategies did not significantly influence motivation. The researchers concluded that self-regulated learners are goal orientated and therefore tend to adopt critical thinking strategies in order to solve difficult tasks and develop skills [10].

Recent research works consider the use of questionnaires to evaluate student motivation in online learning activities. Research reported a strong correlation between learner engagement, motivation, and performance, though such results rely on relatively limited forms of evaluation. The research carried out in this work differs from previous approaches as it employs learning analytics methodologies in analyzing the correlation between learner engagement and performance. Moreover, machine learning is used to identify the lack of motivation in learners, through the discovery of latent patterns of student engagement [9].

D. EXISTING MACHINE LEARNING APPROACHES IN EDUCATIONAL DATA ANALYTICS

Within the educational setting, machine learning is an effective technique that has been widely applied, primarily to the prediction of student performance in both traditional and virtual environments. Kabakchieva [24] applied supervised machine learning methods in predicting student performance at a Bulgarian University. The work considered 20 predictive attributes extracted from personal information and the pre-university characteristics of students. The Bulgarian Score Level scale was used to categorize student performance into five classes, i.e., "Excellent", "Very Good", "Good", "Average", and "Bad". Several supervised ML techniques were used to predict student performance, including Decision Trees, Naive Bayes, Bayesian Networks, and k-Nearest Neighbors. The results demonstrated that the utilized classifier models suffer from low performance, exhibiting an average accuracy in the range of 52-67 % [24]. Asif *et al.* employed data mining methods in predicting the performance of undergraduate students at the Engineering University in Pakistan. Similar to [24], five levels of outcomes were considered as targets, for which the GPA was employed as a predictive feature. The results revealed that the Naive Bayes classifier achieved the highest accuracy, with a value of 83% [25].

A technique called Deep Knowledge Tracing (DKT) was introduced in [26]. The authors applied Recurrent Neural Networks (RNN) on the Khan Academy online courses to predict the future performance of students. RNN is a dynamic model with the ability to continuously represent the state of latent knowledge over time, while evaluating the level of student knowledge. A number of variables were considered for the DKT model, including the student's previous knowledge, student clickstream features, latent engagement, factor difficulty associated with each task, and additionally, the duration of tasks taken by the student during the online sessions. The results showed that the RNN model achieves good performance with an AUC value of 0.85 [26].

Various researchers investigated attrition issues within MOOC environments. Kloft *et al.* [26] applied support vector machines to predict the likelihood of learner withdrawal from online courses, considering only click stream features [27]. Although only one feature was used in the predictive model, feature extraction in the time domain was used to derive higher order attributes, such as the number of sessions, the number of videos watched, and the number of coursework page views. The results showed an accuracy improvement of 15% in the early weeks of the courses, with the highest accuracy obtained at the end week of the courses [27].

Al-Shabandar *et al.* investigated factors driving student withdrawal within MOOCs. The study encompassed data of 7,000 learners enrolled in five courses at Harvard University and MIT. Various machine learning algorithms were applied with the highest prediction accuracy of 94% obtained using the Bagged Cart model, followed by neural networks, with an accuracy of 89% [28].

At-risk students were identified in [29] using the Virtual Learning Environment Dataset (VLE) of the Open University. Two sets of features were considered in this study, namely, behavioral attributes and demographic features. The application of machine learning methods indicated that the proportion of at-risk students increased over time. As such, the precision value dramatically increased from 0.50 at the beginning of the course to 0.90 at the end of the course, while the Recall average value was stable in the range of 0.30-0.50.

Most of the existing work uses surveys and questionnaires to evaluate student motivation in online courses [36]. Machine learning was applied in [36] to predict student motivation. Three sets of features were considered. The "unigram" feature represented the main features set. "Linguistic" features only used student comments in post form. When student comments were positive, then the post was classified as motivated, otherwise unmotivated. The third set of features was "Unigram+Ling", which combined the unigram feature with linguistic features. The results of logistic regression demonstrated that "Unigram+Ling" achieved the best performance with values of 73%, and 62% for Accountable and Fantasy courses, respectively [36].

III. MOTIVATION

One of the main shortcomings of existing research is the lack of explanation for the association between motivation and engagement. The majority of studies employ both quantitative and qualitative methods to measure motivation within MOOCs, relying on the analysis of transcripts, interviews, and survey data. Consequently, learner motivation is evaluated from a rather narrow perspective, which does not account for learner interaction patterns within the MOOCs environment.

Two sets of experiments are conducted in this research. In the first experiment, we investigate the link between the level of engagement and performance, considering the geographical location of the learners. Behavioral features are employed to examine the association of engagement level with performance. As behavioral features are represented with continuous variables, statistical techniques are used in their analysis and interpretation. The statistical analysis makes inferences about the successful and failing learner groups in terms of the number of usage videos and reads chapters. To evaluate whether the descriptive results are significant, we use hypothesis testing (Analysis of Covariance). The findings of the first experiment could facilitate educators in gaining insight into the association of behavioural engagement with academic achievement.

In the second set of experiments, the target is to identify learner motivational status and the reasons behind student drop out from varying viewpoints. Traditional statistical analysis has limited ability in predicting student motivational status as it is not designed to discover the non-linear features that separate the students' motivational categories. Moreover, statistical analysis requires human input in making assumptions about the relationships between variables. Therefore, additional analysis was performed using machine learning techniques that do not rely on classical assumptions. Machine learning approaches are used to categorise learner motivation using predictors extracted from the log data, allowing the interaction of learners to be evaluated. Machine learning is adopted due to its capabilities in analysing high dimensional log data, of arbitrary form, characterized by both noise and complex non-linear pattern components. In the context of the present work, machine learning can be used to identify the lack of learner motivation. Moreover, machine learning enables the analysis of arbitrary forms of correlation between behavioural and demographic features within the online course environment.

IV. RESEARCH METHODOLOGY

A. DATA DESCRIPTION

The dataset used in this study was obtained from Harvard University. Harvard University, in collaboration with Massachusetts Institute of Technology (MIT), pioneer and develop MOOCs. The database comprises 17 courses undertaken through the edX platform, during the first year of their delivery. Across all courses, 597,692 participants were registered, of which only 43,196 users achieved certification. However, around half of the participants never engaged with the courses [30]. The learning material is delivered through a sequence of video lectures, in addition to courseware chapters and a set of quizzes.

In this study, two courses were selected for analysis, namely, "Introduction to Computer Science" and "Circuits and Electronics". The dataset includes some rows with empty values, which were removed in the experiments as part of the data cleaning process.

1) FEATURES

All database features are selected in this study, as shown in Table 1. Harvard University proposed the use of these variables based on a self-reported survey [31]. They delivered the survey to participants encompassing various types of questions regarding the features that potentially impact on learning outcomes. Moreover, they followed the findings of previous research to determine the factors that influence student retention in online courses and determined the parameters that should be taken under consideration, such as student activity and demographics.

Three sets of features are considered in the Harvard dataset, including behavioral (6 features), demographic (5 features), and temporal (2 features), in addition to the user id [30]. The data representation was therefore encoded as a series of vectors.

As shown in Table 1, behavioral features comprised the variables"Nevent", "nplay_video", "Nchapters", "nforum_post", which are integer variables representing discrete counts for each attribute, while "explored" and "viewed" are binary behavioral variables. The "explored" variable is encoded as 1 when a user accessed more than half of the courseware chapters and 0, otherwise. When learners access the courseware home page, including the problem and video sets, the value of "viewed" is set to 1 and 0, otherwise.

TABLE 1. Harvard dataset description.

Features	Type	Description
User-Id	Demographic	Learner identification number
YOB	Demographic	Learner date of birth
Gender	Demographic	Learner gender
LOE	Demographic	Learner educational level
final_cc_cname_DI	Demographic	Learner continent area
Start_time_DI	Temporal	First date learner actvity
last_event_DI	Temporal	Last date learner actvity
ndays_act	Temporal	Number of unique days that the learner interacted with the course
Nevent	Behavioral	Number of click stream events
nplay_video	Behavioral	Number of videos viewed by learner
Nchapters	Behavioral	Number of chapters read by learner
nforum_post	Behavioral	Number of forum postings by learner
Viewed	Behavioral	user access to home page of videos
Explored	Behavioral	user access to home page of chapters

The "educational background" is a demographic parameter, which includes the level of education and consists of a number between 1 and 5, selected from the set of {"less than secondary", "secondary", "bachelors", "masters", "doctorate"}, respectively. The variable "Gender" is given as a categorical demographic variable. The variable, "YOB" stands for the Year of Birth. The variable "final_cc_cname_DI" represents the student geographical area, and is taken from the set of {"Africa", "Asia", "Australia", "America", "Europe"}. The temporal domain raw fields include {"Launch date", "Wrap date", "start_time_DI", "last_event_DI"}. Variable "Launch date" represents the course start date, while variable "Wrap date" represents the issue date of the certification. Variable "start_time_DI" represents the participant's enrollment date, while variable "last_event_DI" is defined as the date of last student activity interaction with the courseware.

2) TARGET CLASSES

As previously mentioned, two courses were selected for the analysis, i.e., "Introduction to Computer Science" and "Circuits and Electronics". In the former, the course focuses on teaching students the use of computation in task solving [31]. The latter course is an introduction to lumped

TABLE 2. Course acronym.

Course	Course Acronym
Circuits and Electronics Fall	Electronics Fall
Circuits and Electronics Spring	Electronics Spring
Introduction to Computer Science and Programming Fall	Computer Fall
Introduction to Computer Science and Programming Spring	Computer Spring

circuit abstraction. The course is designed to serve undergraduate students at Massachusetts Institute of Technology and is available online to learners worldwide [32]. The two courses are selected in our analysis to examine the level of engagement and intrinsic motivation for foundation students.

Fall courses were delivered in the fall of 2012 and spring courses were covered in the spring of 2013. The courses are entitled: "Circuits and Electronics Fall", "Circuits and Electronics Spring", "Introduction to Computer Science and Programming Fall", "Introduction to Computer Science and Programming Spring" as shown in Table 2 [30].

All four courses were ran over a 15 weeks period, including a final exam and two midterm examination periods. There were approximately 150 videos and 14 chapters released in each course. To earn certification, learners must gain a mark above 40% in their overall grade. The overall average grade is calculated from course components, including quizzes (10%), weekly courseware set (40%), two mid term exams (25%), and final exam (25%).

The certification is considered to be an inaccurate indicator of learning within MOOCs [31]–[33]. Due to free enrollment, a large number of learners interact with the course without aiming to undertake the final exam. Moreover, participants who register after the course end date are precluded from obtaining a certificate. However, certificates are a good indicator of learning outcomes for registrants who persisted in completing the course [30].

A data driven approach was employed in this study to categorize learners. The algorithm describes the taxonomy of learners. It relies on Incentive Motivation Theory (IM), where the following categories are defined:

Let V represent a set of students records, where |V| = N is the number of students.

Let $R_i \in V$ represent the i^{th} student record, given as:

Where

v_i - the identity of the student for the i^{th} record

g_i - the grade for the i^{th} student record

s_i - the start date of the associated student with respect to the course

e_i - the end date of the associated student with respect to the course

c_i - the identity of the course associated with the i^{th} entry

l_i - the launch date of the course referred to by c_i

w_i - the wrap date of the certification is issued by c_i

d_i - the number of videos viewed by the i^{th} student

u_i - the number of chapters read by the i^{th} student

Let us consider the retention, completion and attrition learner groups defined as:

Retention Learners (intrinsically motivated)**:** defined as those who engage in a given course without aiming to earn certification as defined in Equation 1:

$$RL = \{\forall v \in V | g = 0 \wedge [[l < s] \vee [w < e]]\} \quad (1)$$

where V is the student records, g is the grade, s is the course start day, l is the course launch day, w is the course wrap date and e is the course end day.

Completion Learners (extrinsically motivated): undertake courses with the expectation of obtaining certification. The group is further categorized in two subsets, i.e., learners who pass and achieve certification, and learners who do not pass. Pass completion learners are defined in Equation 2, while Failure completion learners are defined in Equation 3.

$$CLsc = \{\forall v \in V | g \geq 40 \wedge [[s \leq l] \wedge [e \leq w]]\} \quad (2)$$

$$CLsn = \{\forall v \in V | 0 < g < 40 \wedge [[s \leq l] \wedge [e \leq w]]\} \quad (3)$$

Attrition Learners (Amotivation): defined as learners who withdrew from the course within the same day as expressed in Equation 4.

$$AL = \{\forall v \in V | g = 0 \wedge s = e\} \quad (4)$$

Algorithm 1 shows the groups of learners according to IM Theory. Three groups were defined by considering the students' exam grades, course start and end dates, in addition to the first and last date that students interacted with course. In both the RL and AL groups, students did not undertake the assignment; however, in the RL group, they engage in the course longer than the AL group. Completion learners can be further classified into {CLsc, CLsn}. The assignment cutoff grade was used for distinguishing between these two groups.

Algorithm 1 Learners Group

1. $\forall V \in R^P, \exists R_i : R_i =< v_i, g_i, s_i, e_i, c_i, l_i, w_i, d_i, u_i >$
2. $R_i \in RLs \leftrightarrow g_i = 0; l_i < s_i, w_i < e_i$
3. $R_i \in \leftrightarrow g_i = 0; s_i = e_i$
4. $R_i \in CLsc \leftrightarrow g_i \geq 40; s_i \leq l_i, e_i \leq w_i$
5. $R_i \in CLsn \leftrightarrow 0 < g_i < 40; s_i \leq l_i, e_i \leq w_i$

B. DATA PRE-PROCESSING

Due to the large size of the dataset, a sample of 7000 log file entries was used in each experiment. The log file records represent completed activities undertaken by learners on the respective MOOCs platforms, where each entry corresponds to a single user session.

The data pre-processing is divided into two distinct phases, namely, data cleaning, and data transformation. Data cleaning was used to remove missing values, reduce noise, and remove inconsistencies in the data. On inspection, approximately 15% of the observations were missing for several behavioral variables, namely, Nevent, nplay_video, Nchapters and nforum_post. The YOB, Gender and LoE_DI attributes also included missing values. As a result, all incomplete observations were excluded from the dataset. Moreover, duplicate rows in the dataset were also removed.

The Harvard dataset features have skewed distributions. Consequently, the data could suffer from the presence of non-normality. To overcome this issue, the Box-Cox transformation was used. This is a member of the class of power transform functions, which are used for the efficient conversion of variables to a form of normality, e.g., the equalisation of variance, and to enhance the validity of tests for linearly correlated variables [34]. The data was furthermore processed via scaling and centering such that a mean value of 0 and a standard deviation of 1 were obtained.

C. FIRST SET OF EXPERIMENTS

Various statistical methods were employed in this research to understand the behavioral patterns of learners and explore how behavioral engagement can influence performance in MOOCs courses. Statistical analysis is capable of tracing and tracking learning activities in online courses enabling course designers to gain insight into learners' success and failure within MOOCs platforms. A brief description of the statistical methods explored in our experiments is provided below.

Descriptive statistics: This considers the utilization of the mean and the standard deviation method (μ, σ). These parameters are used in the first set of experiments to compare successful completion learners and unsuccessful completion learners in terms of geographical location and engagement level. Students were distributed in five geographical areas, and two behavioral features were considered, namely, nplay_video and Nchapters. Learners were allowed to reattempt activities frequently; i.e., there was no limit on the number of recorded attempts for each student per activity. Therefore, it was not possible to set a threshold for the number of click events for users watching the videos, and reading pdf files. Descriptive statistics may assist educators to identify the reasons behind student success and failure. Descriptive statistics are defined as follows [35].

$$\mu_j = \left(\frac{1}{N_j} \sum_{i=1}^{N_j} X_{ji} \right) \tag{5}$$

$$\sigma_j = \sqrt{\frac{1}{N_j} \sum\nolimits_{i=1}^{N_j} \left(X_{ji} - \mu_j\right)^2} \tag{6}$$

where j is the location parameter, Nj is the total number of students at location j, and Xji is a student access to an online course from location j.

Analysis of Covariance: To evaluate the results of descriptive statistics, analysis of covariance (ANCOVA) was used. ANCOVA is a statistical test used to test the mean of the independent variable across two groups. In this experiment, ANCOVA was used to determine whether the mean values μ of successful and failing learners are identical with respect to geographical location and engagement level. The ANCOVA variable is defined as [36]:

$$\Upsilon_j = \sum\nolimits_j^m \mu + T_j + \beta(C_j - X\overline{C}_j) + \epsilon_j \tag{7}$$

where m is the number of geographical locations $\{G_1, \ldots, G_m\}$ and n is the number of successful and failing students. In this case, μ is the population mean and $\overline{\mathrm{C}}_{\mathrm{j}}$ refers to the group mean. T_j is the effect of the j^{th} geographical location on the independent variable and ϵ_j is the error term per j^{th} location. β is the slope of regression line and X is the observation under the jth group. C_j is the covariate values of success and failing students' $\mathrm{S_i}$ in the j^{th} geographical location. The $\rho\left(S_i \in G_j\right)$ is the probability of student $\mathrm{S_i}$ belonging to particular geographical area.

The parameter C_j is defined according to the following equation as:

$$C_j = \#\{Students\ S_1, .., S_n \in G_j\} = \sum\nolimits_{i=1}^{n} p\left(S_i \in G_j\right) \tag{8}$$

where $p\left(S_i \in G_j\right)$ is the probability of student $\mathrm{S_i}$ belonging to a particular geographical area G_j.

Chi-squared test: The Chi-squared test is a statistical hypothesis test which was used to examine the difference between failure and success groups per course with respect to the learners' academic level. The Chi-squared test summarizes differences between observed frequency values and expected frequency values for each educational level. The results of the Chi-squared test may help educators in determining whether the educational level factor can impact on learner performance. The Chi-squared test is defined below [37].

$$\chi_j^2 = \sum_j^r \frac{\left(O_j - E_j\right)^2}{E_j} \tag{9}$$

Let r represent the levels of educational background and n represent the total number of successful and failing students: where O_j is the number of successful and failing students per j^{th} educational level described as [38]:

$$O_j = \#\{Students\ S_1, .., S_n\} \in L_j = \sum\nolimits_{i=1}^{n} p\left(S_i \in L_j\right) \tag{10}$$

E_j is the expected frequency of the number of successful and failing students per j^{th} educational level and $p\left(S_i \in L_j\right)$ is the probability of student $\mathrm{S_i}$ belonging to the j^{th} educational level E_j is given by [38]:

$$E_j = \# \sum\nolimits_{i=1}^{n} E(S_i \in L_j) = \sum\nolimits_{i=1}^{n} p(S_i \in L_j) \tag{11}$$

1) ENGAGEMENT LEVEL BETWEEN SUCCESSFUL AND FAILING LEARNERS

Descriptive statistics are computed and stratified according to the demographic region. The engagement levels of learning

TABLE 3. Descriptive statistics analysis: failing learners.

Courses	Mean					SD				
"2012 Courses"	Africa	Asia	Australia	America	Europe	Africa	Asia	Australia	America	Europe
Electronics Spring										
nplay_video	504.077	153.25	114.34	223.64	81.23	663.54	33.65	164.12	375.71	303.10
Nchapters	6.55	5.49	5.26	5.8333	6.27	4.12	3.52	3.89	3.70	4.02
Computer Fall										
nplay_video	213.59	184.04	162.11	76.167	202.78	316.23	286.23	287.63	282.96	267.50
Nchapters	4 .97	5.22	4.67	4.68	5.33	3.57	3.58	3.48	3.20	3.58
"2013 Courses"										
Electronics Spring										
nplay_video	231.17	l 144.25	247.21	209.72	220.53	575.10	256.97	202.88	312.27	3315.21
Nchapters	6.36	5 5.90	6.81	6.42	5.92	4.91	3.60	4.47	3.86	3.41
Computer Spring										
nplay_video	123.13	l 134.20	105.14	130.83	140.05	174.90	266.48	173.74	203.19	217.06
Nchapters	5.08	5 5.21	4.64	4.81	5.27	3.40	3.53	3.09	3.27	3.46

TABLE 4. Descriptive statistics analysis: successful learners.

Courses	Mean					SD				
"2012 Courses"	Africa	Asia	Australia	America	Europe	Africa	Asia	Australia	America	Europe
Electronics Fall										
nplay_video	1304 .6	411.76	729.11	862.67	1177.83	1321.5	711.16	1148.48	912.27	1342
Nchapters	16.34	15.16	14.3	16.30	16.42	2.09	2.12	3.62	1.63	1.64
Computer Fall										
nplay_video	538.64	499.78	720.20	634.12	734.74	579.90	759.37	929.42	509.38	753.15
Nchapters	16.41	16.116	16.36	16.94	17.11	2.34	2.62	2.54	1.69	1.61
"2013 Courses"										
Electronics Spring										
nplay_video	616.5:	333.50	212.66	801.70	1010.67	704.34	505.91	328.35	609.45	1258.61
Nchapters	17.61	16.01	16.43	17.94	17.35	2.25	2.80	2	2.22	2.33
Computer Science Spring										
nplay_video	287.14	319.96	342.6	472.37	560.85	258	458.80	224.29	410.11	567.89
Nchapters	16.6	16.63	16.93	16.54	16.83	1.40	1.46	1.334	1.56	1.51

activities are determined. A comparison of failing groups with successful groups was conducted, accounting for the demographic features of {"Africa", "Asia"," Australia"," America"," Europe"}, in the context of the behavioral features of {nplay_video, Nchapters}. As previously described in table 1 the "nplay_video" represented the number of videos watched by learners and "Nchapters" are the number of chapters read by students.

The results in Tables 3 and 4 indicate that there is a significant difference between the two groups for each course.

The results also demonstrate that successful learners watch more videos than failing students. Europe dominated the top rankings in the successful learners group with $\mu = \{734.74; 1010.67; 560.85\}$ for "Computer Science Fall", "Electronics Spring", and the "Electronics Spring" courses, respectively. However, the highest number of successful learners in "Electronics Fall" lived in Africa with $\mu = 1304.6$.

The results also demonstrate that "Electronics Fall" is the most watched course with approximately 60% of the videos viewed by certified students. Conversely, "Computer Science Spring" was the lowest viewed course, where successful learners viewed only 30% of the videos. Within the successful group of learners, European students watched an average of 42%-51% of videos in both courses, in contrast to the Australian and African counterparts, who viewed the lowest percentage of videos. In the "Computer Science Fall" and "Electronics Spring" courses.

The European learners undertook once again the highest percentage of videos usage, with approximately 50% of the video resources used, and conversely only 1-2% of videos viewed by African and Australian learners. Considering the failing group of students. The largest proportion of videos were watched by Asian participants, who used an average 14% of the video resources in both the "Electronics Fall" and "Computer Science Fall" courses. In the "Electronics Spring" and "Computer Science Spring" courses, American students used around 23% of the videos. In the four courses, the lowest rate of video usage was reported again for Australian participants. The results indicate a significant variability between successful participants and failing learners in respect to the number of chapters read. In general, successful learners read learning materials three times more than unsuccessful learners.

For example, in America, the mean number of chapters read is reported as $\mu = \{16.30; 16.94; 17.94; 16.54\}$ in "Electronics Fall", "Computer Science Fall", "Electronics Spring" and "Computer Science Spring" courses, respectively, for the successful group, in contrast to a reduction by approximately a third in the failing group peers, where $\mu = \{5.38; 4.68; 6.42; 4.81\}$, respectively .

In regards to the "Electronics" courses, the most successful students were reported as Asian, who read 50% of the available learning resources. Europe achieved the highest successful reading activity, with 46% of chapters viewed by the group in the "Computer Science" courses. On average, the percentage of students in the failure group who view course chapters was 70% for the "Electronics Fall" and 66% for the "Electronics Spring" courses, respectively.

Participants within the failing group read only a small proportion of the available course material. Moreover, the proportion of failing students who engaged in reading chapters rose to 90% in the "Computer Science Spring" course, for which the percentage of reading material was slightly higher than in the other courses. For example, approximately 2-20% of the course documents were read by European and African students in the "Electronics Fall" and "Electronics Spring" courses, whereas an average of 16-22% of chapters were read by these learners in the "Computer Science Spring" course.

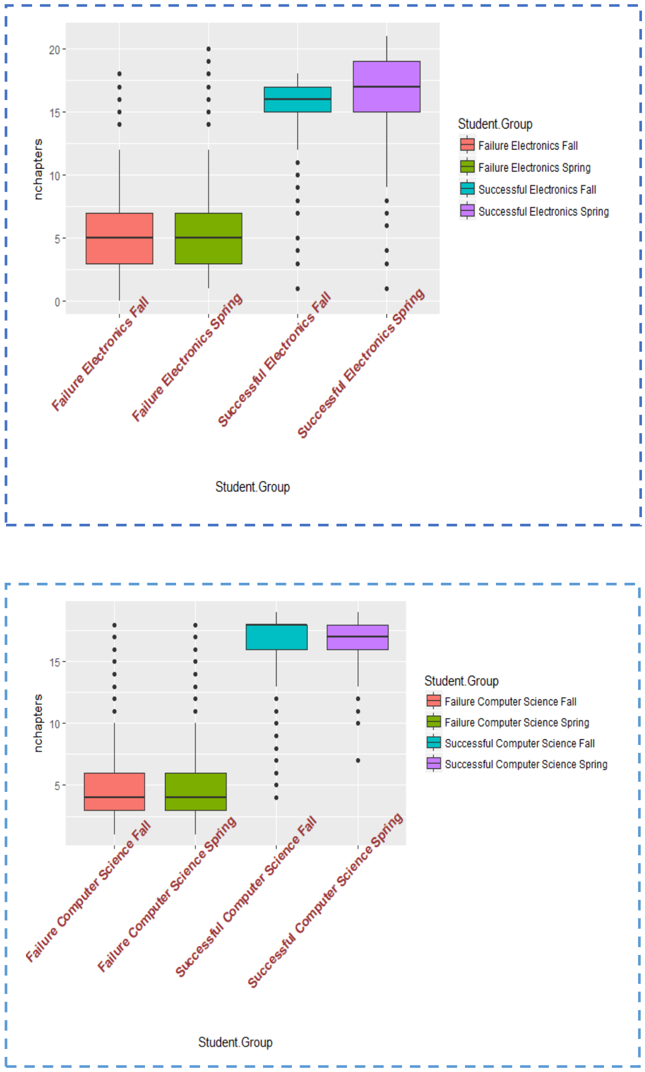


FIGURE 1. **Box plots of failing and successful learners per chapter read.**

In general, the engagement level of the successful group is higher than the failing group, when considering the "nplay_video" and "Nchapters" parameters. Figures 1 and 2 show the box plots in respect to the number of chapters read and videos watched, respectively. The box plots show that the majority of successful learners in the springs courses prefer reading course chapters rather than viewing videos. The number of videos viewed by the successful group is slightly higher in the "Computer Science Fall" course rather than the "Computer Science Spring" course. However, the percentage of reading document is similar in both courses. In this study, ANCOVA is used to determine whether the mean of successful and failure learners are identical regarding engagement level. The result reveals a notable difference between two groups across all courses. The p-value was ($p<0.0002$) for all behavioral features. Hence, there is a significant difference between certified versus failure.

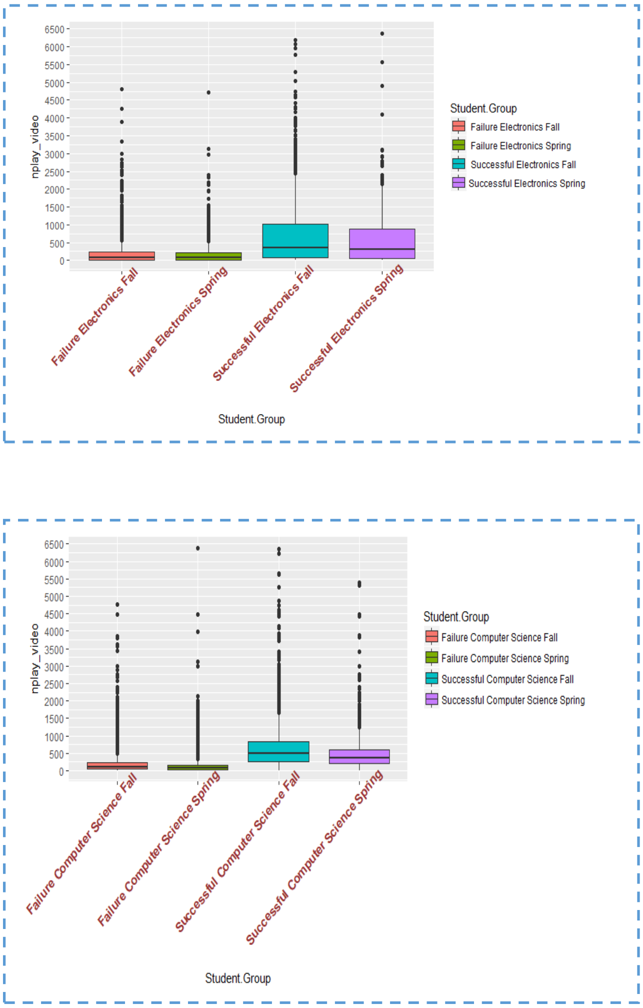


FIGURE 2. Box plots of failing and successful learners per videos viewed.

The following figures shows the distribution of the success and failure of learners per each course in respect to their engagement levels.

2) EDUCATIONAL LEVEL OF FAILING AND SUCCESSFUL LEARNERS

In this section, the association between academic qualifications, demographic features and learner performance is studied. Table 5 illustrates the Chi-squared results. The parameter df stands for the degrees of freedom and can be defined as the number of independent values that vary in the final calculation. The results indicate a p-value of ($p < 0.05$) for all courses except the "Electronics Spring" course, thus allowing for the rejection of the null hypothesis and demonstrating that the learners' educational background is associated with the learners' performance level.

Figures 3 and 4 show the distribution of successful and failing learners for each of the courses with respect to their educational level. Overall, most completion learners are reported as secondary, Bachelors and Masters qualified, with a smaller number of doctorate learners aiming to earn certification.

TABLE 5. Results of the Chi-squared Test comparing failing vs successful learners by educational level.

Course	χ^2 statistic	df	P-value
Electronics Fall	32.012	4	1.902e-06
Electronics Spring	3.4134	4	0.4912
Computer Science Fall	34.734	4	5.268e-07
Computer Science Spring	64.434	4	3.386e-13

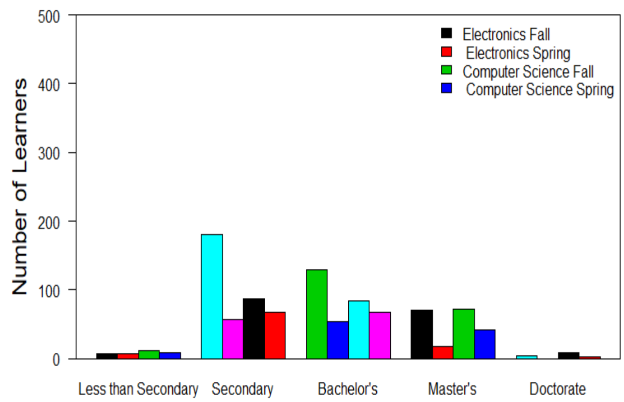


FIGURE 3. Successful learners by educational level.

An average of 40% of learners who have Bachelors or secondary degrees failed in the "Electronics Fall", "Computer Fall", and "Computer Spring" courses. Around 50% of certified learners in the "Electronics Fall" course have a secondary degree, while the percentage of such learners drops to 30%-35% in the "Computer Spring" and "Computer Fall" courses.

Most successful learners with a Bachelors degree are shown in the "Electronics Fall" course. Figures 4 and 5 also show that learners with less than secondary and doctorate qualifications reported the lowest percentage of participation across all courses. An average of 2% of students with less than secondary degrees failed in the "Electronics Fall" and "Electronics Spring" courses, while conversely, the percentage of failing students in the "Computer Science Fall" and "Computer Science Spring" courses is 2% higher, with doctoral qualifications applicable to approximately 0.5%-2% of the student participants. The Chi-squared test and associated Figures demonstrate that the learners' educational level impacts on their performance. In Table 5, the proportion of successful students who have a Masters degree is reported to be around 25%-30% in the "Computer Science Spring" and "Computer Science Fall" courses, while the percentage of Masters qualified learners drops to 18% for the failing groups in both courses.

3) EXPERIMENT DISCUSSION

An empirical comparison between failing and successful learner groups in the first experiment reveals that both demographic and behavioral features could significantly impact on learner performance in an online course. The results of

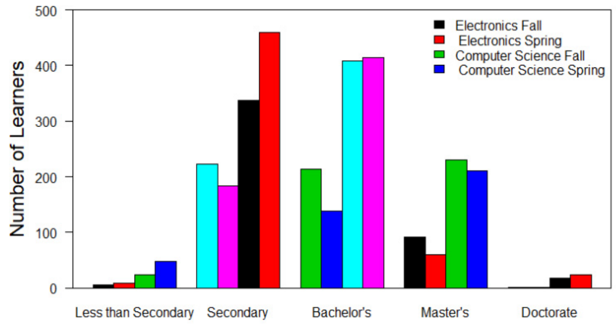


FIGURE 4. Failing learners by educational level.

descriptive statistical analysis show that Europe ranks the highest in terms of learner success rates, while Asia reports the highest ratio of failing group participants. Due to the lack of advanced technological integration within universities and colleges in Asia, students in these regions are likely to face technical issues. Additionally, the language of instruction is considered as another barrier, since courses are delivered in English, hence learners may be less motivated to exchange knowledge with other participants.

Accordingly, analysis of such descriptive statistics could assist educators and course instructors in enhancing learning resources by early identification of at risk students. Algorithm 2 shows the proposed statistical analysis process for the separation of CLsn and CLsc learners, while Figure 5 shows the flow chart of the proposed algorithm. The mean (μ) andStandard deviation (SD) are computed for each geographical location. In this case, t1 and t2 are statistically significant thershold values, which should be determined according to the learners' behaviors related to a specific region.

The Chi-squared test was applied to investigate the presence of a significant difference in the educational levels of learners in the successful and failing groups. The results suggest that the educational background could be an important factor affecting learner performance in online classes. Masters qualified students show the largest percentage of successful completion. While statistical analysis is informative, it is not designed to capture arbitrary non-linear patterns. As a result, such procedures require expert assumptions about the form of the data prior to analysis, relying on the notion of a super population whose form must be chosen on an a priori basis [39].

Moreover, in the context of our investigation, hypothesis tests and inference procedures are not conducive to the identification of withdrawal students, since the data is not guaranteed to satisfy classical statistical constraints. To understand the reasons behind student withdrawal, important factors affecting learner motivation need to be identified thus leading to the application of advanced learning analytics methods. Advanced analysis was therefore considered using machine learning models that do not based on classical assumptions.

The machine learning is used in our study to help educators flag lack of motivational students at their early stages of the course and deliver timely intervention assistance to those students.in addition, the course instructors could immediately provide support for these students, by improving their motivation and increasing their engagement.

Algorithm 2 Proposed Statistical Analysis for *CLsc* and *CLsn* Learners

1. $\forall V \in R^P, \exists R_i : R_i =<, g_i, s_i, e_i, c_i, l_i, w_i, d_i, u_i>$
2. $R_i \in R \wedge$ i = 1, ..., n, where n is the number of students
3. Let c $\in C$ = MOOCs course and 1 $\in$ L = Geographical location
4. $R_l \in R = \{R_i \wedge R_i \; per \; location \; l\}$
5. Find d_l *per* R_l
6. Calculate (μ_l, σ_l) *per* d_l
7. Find u_l *per* R_l
8. Calculate (μ_i, σ_i) *per* u_l
9. If $(\mu_l\sigma_l) < t_1 \wedge (\mu_j\sigma_j) < t_2$
10. $R_l \in CLsn$
11. else
12. $R_l \in CLsc$
 where t_1 and t_2 are predefined threshold values
13. End If
14. End

D. SECOND SET OF EXPERIMENTS

In this set of experiments, a number of machine learning algorithms are applied in the prediction of learner motivation. The purpose of this investigation is to assist instructors in early detection of lack of participant motivation in MOOCs. Machine learning provides the ability to model and autonomously categorize learners into motivation classes. In this investigation, learner behavior in conjunction with learning outcomes is considered in the classification of learner motivation cues based on IM theory. Multi-class classification is used, where the set of labels, 1,...,L, represents the target classes. Learner motivation is classified into three classes/labels, i.e., intrinsic, extrinsic, and amotivation. The training dataset consists of the pairs $(\boldsymbol{F_i}, \boldsymbol{T_i})$, where $\boldsymbol{F_i} \in \mathbb{R}^{\boldsymbol{p}}$, denotes features of i^{th} observation and $\boldsymbol{T_i}$ are the associated targets, $\boldsymbol{T_i} \in \{1, \ldots, \mathrm{L}\}$.

The confusion matrix was used to evaluate predictive model performance. Furthermore, the sensitivity, specificity, F_1-Measure, and accuracy were used for the purposes of evaluation. Precision or positive predictive value (PPV) is defined as the ratio of true positives (TP) over the total number of positives, P=TP+FP, where FP is the number of false positives. Recall or negative predictive value (NPV) measures the ratio of true negatives (TN) over the total number of negatives, N=TN+FN, where FN is the number of false negatives. The F_1-Measure is used to test the accuracy of the classifier models, accounting for both precision and recall. Specifically, the F_1-score is defined as the harmonic of the

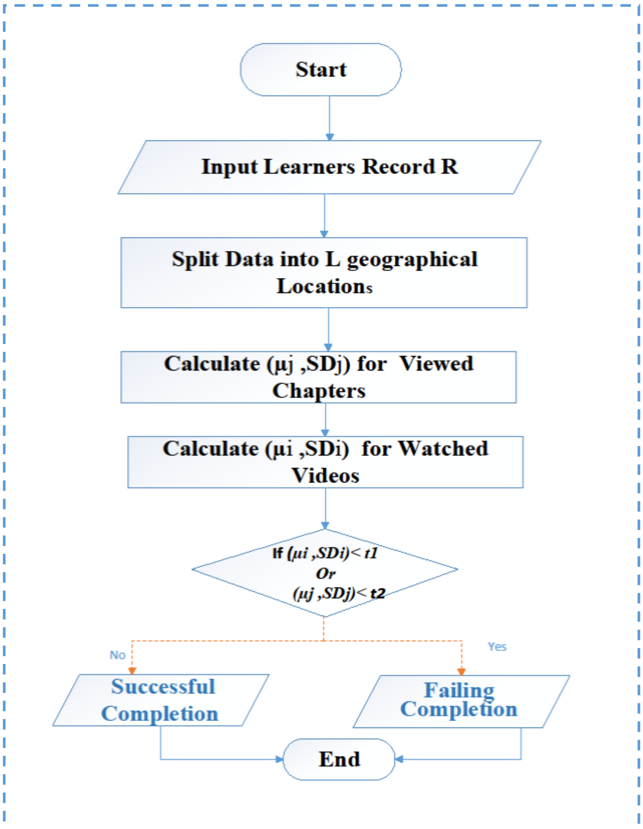


FIGURE 5. Algorithm 2 Flow chart.

precision and recall values [40]. The performance measures are defined as:

Sensitivity = True Positive Rate (TPR)

$$\mathrm{TPR} = p(\hat{C} = \oplus | C = \oplus) \simeq \frac{TP}{P} \tag{12}$$

Specificity = True Negative Rate (TNR)

$$\mathrm{TNR} = p(\hat{C} = \ominus | C = \ominus) \simeq \frac{TN}{N} \tag{13}$$

False Positive Rate (FPR)

$$\mathrm{FPR} = \frac{FP}{FP + TN} \tag{14}$$

False Negative Rate (FNR)

$$\mathrm{FNR} = \frac{FN}{FN + TP} \tag{15}$$

Accuracy (ACC)

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} \tag{16}$$

Precision (PPV)

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{17}$$

Recall (NPV)

$$\mathrm{NPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{18}$$

F_1-score (F_1)

$$F_1 = \frac{2}{\frac{1}{NPV} + \frac{1}{PPV}} \tag{19}$$

The TP, TN, FP, FN, may be derived through appropriate computations applied to the empirical prediction responses and respective correct class values, as given in Equations 12,13,14, and 15.

The Receiver Operator Characteristic (ROC) and Area Under Curve (AUC) metrics are also considered. The ROC is a graphical representation in which TPR is plotted against FPR, producing a parametric curve that may subsequently be used to select appropriate cut-off values. The AUC is defined as:

$$AUC = \int_0^1 \frac{\mathrm{TP}}{\mathrm{P}} d\frac{FP}{N} = \frac{1}{PN} \int_0^N TP\, dFP \tag{20}$$

The AUC is used to measure the predictive quality of a classification model, with the perfect classifier producing a value of 1. A probabilistic classifier randomly assigns scores for positive instances and negative instances [41]. The scoring is computed based on the Mann Wilcoxon test (w) rules. The Mann Wilcoxon test is a nonparametric test used to detect if the observations in two different populations are identical. The w test rules are described as [42]:

$$s(X_p, X_n) = \begin{cases} 1, & \text{if } X_p > X_n \\ 0.5, & \text{if } X_p = X_n \\ 0, & \text{if } X_p < X_n \end{cases} \tag{21}$$

The AUC is equivalent to the Mann Wilcoxon test (w) and can be computed as:

$$AUC = w = \frac{1}{PN} \sum\nolimits_{X_p \epsilon pos} \sum\nolimits_{X_n \epsilon neg} s(X_p, X_n) \tag{22}$$

$$AUC = p\left(X_p > X_n\right) + \frac{1}{2} p\left(X_p = X_n\right) \tag{23}$$

where $s(X_p, X_n)$ is the score for the probabilistic classifier, and X_p, X_n are probability rankings of examples that belong to the positive and negative classes, respectively.

1) MODEL CONSTRUCTION AND VALIDATION

A ten-fold cross-validation is applied during the modelling, with five repetitions. The 50% of the original dataset was allocated for cross-validation. For each round of cross-validation, nine folds subset are used to train classifiers and one is used as a test sample. The training set consists of a total of 4,060 data points which were randomly sampled from the subset of the courses considered, namely, "Electronics Spring" and "Computer Science Spring".

A further 50% of the data, 4000 records, disjoint from the cross-validation set and it is used as an external test dataset to validate generalization errors for each classifier. These data points were randomly sampled from a separate subset of courses, comprising "Electronics Fall" and "Computer Science Fall". The classifier models were trained using the data from one set of courses (i.e., Spring), and the learned predictive models were tested on a previously unseen set of courses (i.e., Fall).

This process supports the investigation of the generalization of the associated mappings, learned by the classifiers, to be examined beyond the specifics of an individual set of courses. The percentages of patterns for cross-validations and test sets from the intrinsic, extrinsic and amotivation classes are 29%, 29% and 42%, respectively.

2) MACHINE LEARNING ALGORITHMS UTILIZED IN THIS EXPERIMENT

In this section, we provide a brief overview of the main machine learning techniques utilized in the present work.

a: DECISION TREE

A decision tree is a hierarchical subtype of the directed acyclic graph (DAG), constrained by performing two steps, recursion and partitioning. The tree structure consists of three canonical components: a root node, a set of internal nodes, and a set of leaf nodes. Each node acts as a processing element that acts on a subset of the pattern space, performing a logical test on a particular attribute, for which outcomes are propagated by outgoing edges [43]. Each successive transfer from a parent to a child node is adapted such that the homogeneity of the resulting pattern is increased with respect to the outcome classes, a property defined as purity. Attributes of the highest discriminative power are represented in the root node. With lessening power towards the leaf nodes, the overall objective is that all leaf nodes will be completely pure.

The main advantage of the decision tree is that the output can be easily interpreted, even by non-specialists, as it is represented in graphical form [44]. Another benefit is in handling nominal and numeric parameters; it is a nonparametric method that does not require normalization of data. In addition, the decision tree can handle databases that have missing and error values. As a consequence, it could easy to incorporate with other classification approaches [44], [45].

One of the main drawbacks of the decision tree is the overfitting phenomenon. As mentioned, the concept of creating a decision tree model depends on a split dataset, which leads to increasing the number of nodes and reducing the number of training rate errors [46].

Let X_t represent a set of training examples relevant to node t and Y=$\{Y_1, \ldots, Y_c\}$ is a set of target classes. The tree is constructed by spliting the observation feature X into the various groups. For continuous features, the tree is grown up based on a set of test conditions and questions with expected results in terms of binary outcomes, i.e., {yes, no}. Node t is partitioned into two branches as follows:

$$\begin{aligned} \mathrm{t}_l &= \{\mathrm{t} \in X\mathrm{:A} \leq \mathrm{V}\} \\ t_r &= \{\mathrm{t} \in X : \mathrm{A} > \mathrm{V}\} \end{aligned} \tag{24}$$

where A is the test condition with outcome V $\in \{0, 1\}$, t_l *and* t_r represent the left and right nodes of the new tree t. To evaluate the best split in feature space, a variety of measures have been utilised including Entropy, Gini, and classification error defined as follows [46].

$$\mathrm{Entropy(t)} = \sum\nolimits_{i=0}^{C-1} p(i|t)\log_2 p(i|t)) \tag{25}$$

$$\mathrm{Gini(t)} = 1 - \sum\nolimits_{i=0}^{C-1} [p(i|t)]^2 \tag{26}$$

$$\mathrm{Classification\ error(t)} = 1 - \max_i[p(i|t)] \tag{27}$$

where $p\,(i\,|\,t)$ is the probability of recodes associated with class i at a given node t and C is the number of classes.

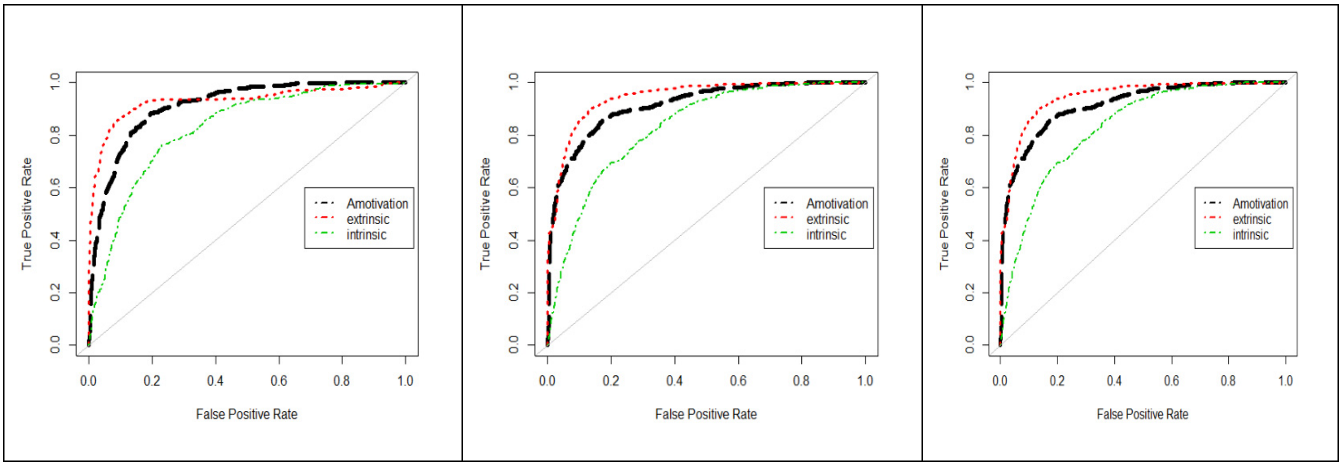
b: NEURAL NETWORKS

Neural Networks are a problem solving methodology grounded in the connectionist paradigm, comprising of networks of interconnected elementary units whose adaptive parameters maybe tuned to form an emergent solution. In particular, neural networks are modelled as a canonicalized abstraction of the biological neural networks found in the mammalian brain, aiming to capture the information processing capability of such structures [47]. In Multilayer Perceptrons (MLP), which is a type of feedforward neural networks, information is transferred forward in one direction without cycles. Neurons belonging to layer (i) receive information from neurons in layer $(i - 1)$ and transmit it to neurons in layer $(i + 1)$, and so on. The input units are connected to the output layer through a sequence of weighted edges.

During the training process, the backpropagation algorithm is typically used to compute and adjust weights in response to some error signal, given some input features [47]. The backpropagation algorithm is employed to compute neural network weight through gradient descent. More specifically, the optimization algorithm, gradient descent is utilized to update the weights of the network by computing the gradient of the loss function. In the context of learning, the cost function computes the error between the actual inputs and the predicted outputs then calculated errors are propagated backwards to the previous layer. During the learning the weight is adjusted iteratively via the gradient descent algorithm until the errors of cost function is minimized [48].

Neural networks can learn to model complex relationships between features; therefore, it has been used to find accurate solutions of complex problems that are difficult to solve by humans or through traditional methods. Another advantage of neural networks is that they can quickly make the correct prediction on unseen data. The new data can generalized even it has high levels of noise [49].

In the context of the present study, behavioral features are used in conjunction with demographic features, each corresponding to a node in the input layer of the neural network. The output layer contains three nodes which represent each class of student motivation, respectively, such that the network can be formally defined as: W, the set $\{(W^i)\}_{i=1}^{N-1}$, denotes the weight matrix connecting layers i and $i + 1$ for a network of N layers, and, B is the set $\{(B^i)\}_{i=1}^{N-1}$, where B^i denotes the column vector of biases for layer i.



(a) (b) (c)

FIGURE 6. Roc curves. (A) DT model. (B) NN model. (C) RDA model.

c: REGULARIZED DISCRIMINANT ANALYSIS

Fisher's linear discriminant used in classification problems, characterized by high dimensional data and small sample size. Shrunken centroids RDA [50] is a generalization of RDA, which is capable of eliminating the overfitting of data by setting an optimal threshold. Shrunken centroids RDA provides a tradeoff between linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA).

In LDA, a common covariance matrix is assumed for all classes, while in QDA, individual covariance matrices are assigned to each class. RDA shrinks the individual covariance matrix for each class to a common covariance matrix [50]. RDA is more appropriate in classification of high dimensional data than LDA.

RDA is efficient technique. It can be easily implemented as it does not require scaling of the features and choice of tuning parameters. The advantage of logistic regression is that the cost with respect to computational complexity is low. A critical limitation of logistic regression, however is that it is unable to solve nonlinear problems since it is a generalized linear model [50].

3) SECOND SET OF EXPERIMENTAL RESULTS AND DISCUSSION

The empirical results were compared using the performance metrics previously described. Each set of analyses were ran five times. Mean average values were yielded over 5 rounds of simulations for the F1-Measure, Precision, Recall and AUC. Tables 6 and 7 summarize the overall accuracy and kappa results, showing the best result of 0.7546 generated by the DT network, while the lowest result was achieved by RDA, with an average value of 0.7372. Table 7 provides an overview of the results.

Actual values refer to the ground truth values of each class over the test dataset, with predicted values referring to predicted classes (motivation category), as obtained from

TABLE 6. Empirical result for classifier model accuracy and kappa.

Classifier	Accuracy	Kappa
Decision Tree	**0.7546**	**0.8486**
Neural Networks	0.7376	0.6323
Regularized Discriminant Analysis	0.7372	0.6092

each classifier model. For example, the DT model predicts 474 learners as intrinsic out of 767 actual values for class "intrinsic". The classifier correctly predicts 501 out of 618 learners as belonging to the "amotivation" class. The highest correct prediction is reported in the "extrinsic" class, where 544 learners are correctly classified out of 628. Class "extrinsic" yielded the highest precision (true positive ratio), with range values of 86%-88% for all classifier models. Class "intrinsic" had the highest recall (true negative ratio) for RDA and NN. DT provided the best specificity results for Class "amotivation".

Table 7 shows that recall is higher than precision across all classifiers however, in NN and RDA, precision gives better results for class "amotivation" with average values of 86%-88%. DT achieved strong precision results for the "amotivation" class, yielding a value of 0.89%. Moreover, there is no noticeable difference between class precision for all models; the "amotivation" "class obtained higher precision than the other classes achieving values of 0.88 and 0.86, respectively. Conversely, analysis of the "intrinsic" class gives a lower precision over all models with average values of 0.50-0.61.

We used ROC analysis to select a decision threshold value for the true and false positive rates. Figure 6 shows the

TABLE 7. Classifier prediction performance results.

Classifier Decision Tree		Actual Amotivation	Actual extrinsic	Actual intrinsic	Precision	Recall	F1-Measure	AUC
Predicted	Amotivation	501	9	144	0.810	0.893	0.790	0.880
	extrinsic	19	544	153	0.866	0.875	0.809	0.895
	intrinsic	98	75	474	0.618	0.861	0.670	0.739

Neural Network		Actual Amotivation	Actual extrinsic	Actual intrinsic	Precision	Recall	F1-Measure	AUC
Predicted	Amotivation	544	23	246	0.883	0.805	0.761	0.911
	extrinsic	13	547	128	0.873	0.897	0.832	0.931
	intrinsic	59	56	385	0.507	0.907	0.611	0.824

Regularized Discriminant Analysis		Actual Amotivation	Actual extrinsic	Actual intrinsic	Precision	Recall	F1-Measure	AUC
Predicted	Amotivation	532	42	201	0.860	0.825	0.763	0.915
	extrinsic	11	553	167	0.880	0.871	0.813	0.946
	intrinsic	75	33	399	0.520	0.913	0.626	0.829

similarity of performance for all classifier models, achieving a range of AUC values between 82%-94% across all classes, however, DT for class ''intrinsic'' provided the lowest AUC. As indicated in Table 7, the F1-Measure for NN shows slightly better results than DT. The lowest F1-Measure is reported for class ''intrinsic'' with a value of 0.6111 for the NN model.

The main reason for DT achieving the highest performance is that it employs operations research principles when predicting the label class based on decision rules [43]. Moreover, it provides an easily accessible representation, which may be used to understand which features impact on prediction.

In our case, we found that the clickstream followed by the ''ndays_act'' features were the most important parameters for prediction purposes. The overall results indicate that there is no major difference between the accuracy of the neural network and Regularized Discriminant Analysis. One possible explanation for the neural network's slightly superior performance, in comparison to Regularized Discriminant Analysis, is the ability to build internal abstractions to aid in the analysis of the complex relationships between the input features and the target [49]. The hidden units, in neural networks, create a new feature space, which can be used to facilitate class discrimination. However, neural networks impede on the explainability of feature contributions.

In general, all the classifier models perform well. The good results demonstrated that behavioral features combined with demographics are capable of distinguishing students' motivational statues. Due to such elements having strong associations with the target class. The results reveal that behavioral features can be used to detect the lack of students' motivation at the early stage of online courses.

V. CONCLUSIONS

The present study conducted two sets of experiments with the aims of providing instructors and course designers with information to assist them in enhancing online courses. In the first experiment, a set of behavioral features was taken into consideration. A descriptive statistical test was used to compare successful students versus failing ones; the results demonstrated a significant difference between the two groups in terms of engagement level. A small number of participants, who used more than half of the learning resources, succeed in all courses. Furthermore, the correlation between the participants' educational level and performance was also examined by conducting a Chi-squared test.

The test outcome rejects the null hypothesis and indicates the presence of a significant difference in variances between the successful and failing groups in terms of engagement level. The test results showed that the educational level is a critical factor impacting on learner performance. In general, around 40% of the participants are educated to either secondary or Bachelors degree level.

In the second set of experiments, machine learning algorithms were applied in the prediction of learner motivation in MOOCs environments. Three classes of motivation were considered in this study, i.e., intrinsic, extrinsic, and amotivation. The best accuracy was achieved using the DT model with a value of 75%, whereas the lowest performing classifiers were NN and RDA, attaining values of 73%. Although all classifers demonstrated approximately similar classification performance, the NN and RDA models obcure the interpretation of factors affecting learner motivation. Our research indicates that DT is a more suitable classifier, achieving a good level of accuracy. In contrast to the other models, the decision tree identifies the click stream and the number of unique days that learners interact with the course as the most important features. Armed with knowledge of the important features, course designers may gain richer understanding of reasons behind learner motivation within the online course setting.

In terms of future work, sentiment analysis can be utilized in interpreting student opinions for learning in MOOCs. As such, the post-forum can be used to capture student attitudes and in Identify those who tend to drop out from the courses. Different emotional statuses can be inferred from discussion forums such as frustration, fatigue, and boredom. These statuses provide the student with motivational encouragement and stimulation to facilitate an interactive learning environment; including feedback modalities, such as visually oriented hints. Additionally, instructors would be able to understand the reasons behind student withdrawal.

REFERENCES

[1] A. D. Ho *et al.*, "HarvardX and MITx: Two years of open online courses fall 2012-summer 2014," *SSRN Electron. J.*, no. 10, pp. 1–37, 2015.
[2] J. Qiu *et al.*, "Modeling and predicting learning behavior in MOOCs," in *Proc. 9th ACM Int. Conf. Web Search Data Mining*, 2016, pp. 93–102.
[3] C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy, "Visualizing patterns of student engagement and performance in MOOCs," in *Proc. 4th Int. Conf. Learn. Anal. Knowl. (LAK)*, vol. 14, 2014, pp. 83–92.
[4] K. F. Hew, "Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS," *Brit. J. Educ. Technol.*, vol. 47, no. 2, pp. 320–341, 2016.
[5] D. Gašević, C. Rose, G. Siemens, A. Wolff, and Z. Zdrahal, "Learning analytics and machine learning," in *Proc. 4th Int. Conf. Learn. Anal. Knowl. (LAK)*, 2014, pp. 287–288.
[6] D. Clow, "The learning analytics cycle: Closing the loop effectively," in *Proc. 2nd Int. Conf. Learn. Anal. Knowl. (LAK)*, 2012, pp. 134–138.
[7] D. F. O. Onah, J. Sinclair, and R. Boyatt, "Dropout rates of massive open online courses: Behavioural patterns," in *Proc. EDULEARN*, 2014, pp. 5825–5834.
[8] F. Wang and L. Chen, "A nonlinear state space model for identifying at-risk students in open online courses," in *Proc. 9th Int. Conf. Educ. Data Mining*, 2016, pp. 527–532.
[9] M. Barak, A. Watted, and H. Haick, "Motivation to learn in massive open online courses: Examining aspects of language and social engagement," *Comput. Educ.*, vol. 94, pp. 49–60, Mar. 2016.
[10] M.-H. Cho and M. L. Heron, "Self-regulated learning: The role of motivation, emotion, and use of learning strategies in students' learning experiences in a self-paced online mathematics course," *Distance Educ.*, vol. 36, no. 1, pp. 80–99, 2015.
[11] M. Hartnett, A. S. George, and J. Dron, "Examining motivation in online distance learning environments: Complex, multifaceted and situation-dependent," *Int. Rev. Res. Open Distrib. Learn.*, vol. 12, no. 6, pp. 20–38, 2011.
[12] V. Trowler, "Student engagement literature review," *Higher Educ. Acad.*, vol. 11, no. 1, pp. 1–15, 2010.
[13] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses," in *Proc. 3rd Int. Conf. Learn. Anal. Knowl.*, 2013, pp. 170–197.
[14] C. Coffrin, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Engaging with massive online courses," in *Proc. 23rd Int. Conf. World Wide Web (WWW)*, vol. 34, no. 6, 2014, pp. 687–698.
[15] K. F. Hew, "Towards a model of engaging online students: Lessons from MOOCs and four policy documents," *Int. J. Inf. Educ. Technol.*, vol. 5, no. 6, pp. 425–431, 2015.
[16] G. Balakrishnan and D. Coetzee, "Predicting student retention in massive open online courses using hidden Markov models," Dept. Elect. Eng. Comput. Sci., Univ. Calif. Berkeley, Berkeley, CA, USA, Tech. Rep., 2013, pp. 1–15.
[17] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé, III, and L. Getoor, "Modeling learner engagement in MOOCs using probabilistic soft logic," in *Proc. NIPS Workshop Data Driven Educ.*, vol. 16, 2013, pp. 1–7.
[18] D. Martimort, "Exclusive dealing, common agency, and multiprincipals incentive theory," *RAND J. Econ.*, vol. 27, no. 1, pp. 1–31, 1996.
[19] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemp. Educ. Psychol.*, vol. 25, no. 1, pp. 54–67, 2000.
[20] J. W. Osborne and B. D. Jones, "Identification with academics and motivation to achieve in school: How the structure of the self influences academic outcomes," *Educ. Psychol. Rev.*, vol. 23, no. 1, pp. 131–158, 2011.
[21] S. M. Glynn, P. Brickman, N. Armstrong, and G. Taasoobshirazi, "Science motivation questionnaire II: Validation with science majors and nonscience majors," *J. Res. Sci. Teach.*, vol. 48, no. 10, pp. 1159–1176, 2011.
[22] P. G. de Barba, G. E. Kennedy, and M. D. Ainley, "The role of students' motivation and participation in predicting performance in a MOOC," *J. Comput. Assist. Learn.*, vol. 32, no. 3, pp. 218–231, 2016.
[23] H. B. Shapiro, C. H. Lee, N. E. W. Roth, K. Li, M. Çetinkaya-Rundel, and D. A. Canelas, "Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers," *Comput. Educ.*, vol. 110, pp. 35–50, Jul. 2017.
[24] D. Kabakchieva, "Predicting student performance by using data mining methods for classification," *Cybern. Inf. Technol.*, vol. 13, no. 1, pp. 61–72, 2013.
[25] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, pp. 177–194, Oct. 2017.
[26] C. Piech *et al.*, "Deep knowledge tracing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 505–513.
[27] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, "Predicting MOOC dropout over weeks using machine learning methods," in *Proc. EMNLP Workshop Anal. Large Scale Social Interact. (MOOCs)*, vol. 4, no. 3, 2014, pp. 60–65.
[28] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, and J. Lunn, "Towards the differentiation of initial and final retention in massive open online courses," in *Proc. Int. Conf. Intell. Comput.*, 2017, pp. 26–36.
[29] J. Kuzilek, M. Hlosta, D. Herrmannova, Z. Zdrahal, and A. Wolff, "OU analyse: Analysing at-risk students at The Open University," *Learn. Anal. Rev.*, no. LAK15-1, pp. 1–6, 2015.
[30] A. D. Ho *et al.*, "HarvardX and MITx: The first year of open online courses, fall 2012-summer 2013," *SSRN Electron. J.*, no. 1, pp. 1–33, 2014.
[31] J. Guttag, "6.00x introduction to computer science and programming MITx on edX—2013 spring," New York, NY, USA, Tech. Rep., 2014.
[32] P. F. Mitros *et al.*, "Teaching electronic circuits online: Lessons from MITx's 6.002x on edX," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2013, pp. 2763–2766.
[33] J. W. D. Seaton, J. Reich, S. Nesterko, and T. Mullaney, "6.00x introduction to computer science and programming MITx on edX—2012 fall," New York, NY, USA, Tech. Rep., 2014.
[34] J. W. Osborne, "Improving your data transformations: Applying the box-cox transformation," *Pract. Assessment, Res. Eval.*, vol. 15, no. 12, pp. 1–9, 2010.
[35] X. Wan, W. Wang, J. Liu, and T. Tong, "Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range," *BMC Med. Res. Methodol.*, vol. 14, no. 1, pp. 1–13, 2014.
[36] P. Gribble, *Analysis of Covariance (ANCOVA) Introduction to Statistics Using R (Psychology 9041B)*, document, 2014, pp. 1–14.
[37] A. Satorra and A. Satorra, "A scaled difference chi-square test statistic for moment structure analysis," *Psychometrika*, vol. 66, no. 4, pp. 507–514, 2001.
[38] A. Agresti, *An Introduction to Categorical Data Analysis*. New York, NY, USA: Wiley, 2007.
[39] C.-T. Chu *et al.*, "Map-reduce for machine learning on multicore," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19, 2007, pp. 281–288.
[40] J. Hastie, T. Tibsharani, and R. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer-Verlag, 2009.
[41] M. Vuk and T. Curk, "ROC curve, lift chart and calibration plot," *Metodoloski Zvezki*, vol. 3, no. 1, pp. 89–108, 2006.
[42] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
[43] R. K. Zimmerman *et al.*, "Classification and regression tree (CART) analysis to predict influenza in primary care patients," *BMC Infectious Diseases*, vol. 16, no. 1, p. 503, 2016.
[44] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, "Decision trees: An overview and their use in medicine," *J. Med. Syst.*, vol. 26, no. 5, pp. 445–463, 2002.
[45] E. M. Rounds, "A combined nonparametric approach to feature selection and binary decision tree design," *Pattern Recognit.*, vol. 12, no. 5, pp. 313–317, 1980.

[46] M. Pal and P. M. Mather, ''An assessment of the effectiveness of decision tree methods for land cover classification,'' *Remote Sens. Environ.*, vol. 86, no. 4, pp. 554–565, 2003.
[47] S. Singhal and L. Wu, ''Training multilayer perceptrons with the extended Kalman algorithm,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 1988, pp. 133–140.
[48] J. A. Bullinaria, ''Learning in multi-layer perceptrons, back-propagation,'' *Neural Comput. Lect.*, vol. 7, no. 8, pp. 1–16, 2015.
[49] P. Perner, U. Zscherpel, and C. Jacobsen, ''A comparison between neural networks and decision trees based on data from industrial radiographic testing,'' *Pattern Recognit. Lett.*, vol. 22, no. 1, pp. 47–54, 2001.
[50] Y. Guo, T. Hastie, and R. Tibshirani, ''Regularized linear discriminant analysis and its application in microarrays,'' *Biostatistics*, vol. 8, no. 1, pp. 86–100, 2006.

RAGHAD AL-SHABANDAR received the B.Sc. degree from the University of Al-Mustansiriya, Iraq, and the M.Sc. degree in information system from Liverpool John Moores, U.K., in 2014, where she is currently pursuing the Ph.D. degree with the Faculty of Engineering and Technology. Her research interests include technology-enhanced learning, machine learning, and classifications.

ABIR JAAFAR HUSSAIN received the Ph.D. degree from The University of Manchester, U.K., in 2000. Her Ph.D. thesis was on Polynomial neural networks for image and signal processing. She is currently a Professor of machine learning and the Head of the Applied Computing Research Group, Faculty of Engineering and Technology, Liverpool John Moores University, U.K. She is also a Ph.D. supervisor and an External Examiner for research degrees, including Ph.D. and M.Phil. She has published numerous referred research papers in conferences and journal in the research areas of e-learning, neural networks, signal prediction, telecommunication fraud detection and image compression.

PANOS LIATSIS received the Diploma degree in electrical engineering from the University of Thrace, Greece, and the Ph.D. degree in electrical engineering from the Control Systems Centre, The University of Manchester. He is a Professor of image processing and the Head of the Information Engineering and Medical Imaging Centre. He is currently the Department Chair of the Electrical Engineering at Petroleum Institution, United Arab Emirates. He has authored/co-authored over 150 research publications in international journals, book chapters, and conference proceedings. He has over 20 years' expertise in the development of advanced sensors, pattern recognition, and intelligent systems, with focus on medical imaging.

ROBERT KEIGHT is currently pursuing the Ph.D. degree with the Faculty of Engineering and Technology, Liverpool John Moores University. His research interests include technology-enhanced learning, machine learning, and e-health.