Learning Analytics for Predicting Student Performance in Online Learning Environments

Sayed Mahbub Hasan Amiri
Department of Computer Science, Dhaka Residential Model College
amiri@drmc.edu.bd

Abstract

The fast-growing numbers of the online learning space have led to the storage of huge amounts of student-to-student interaction data in Learning Management Systems (LMS). However, very often, the educational institutions do not have systematic systems of the usage of such data to help to identify students who are at risk of the future changes in time. This paper fills this gap by building and testing predictive models to predict academic performance of students through learning analytics. Using a quantitative research design, we studied the interaction logs, assessment data, and recorded engagement of 350 university students taking a course all semester-long through Moodle. The most essential behavioral variables, such as the number of logins, the timeliness of submission of assignments, success of discussion forums and watching video lectures were extracted and were used to train and compare various machine learning models, namely, Logistic Regression, Random Forest, and Support Vector Machines. Accuracy, precision, recalls and F1-score were used to measure model performance. Findings indicate that the highest predictive accuracy is experienced in the Random Forest (87-percent), and the assignment submission pattern and a regular frequency of logging into the account are the most potent predictors of ultimate academic achievement. These findings highlight the possibility of learning analytics to support early warning systems based on data, which is why early pedagogical interventions can be provided. This paper becomes a contribution to the literature on educational data mining through the empirical evidence of the relationships between behavioral indicators based on conventional LMS logs and their good predictive abilities of student results, which would provide practical implications to teachers, instructional designers, and institutional policymakers seeking to increase student learning and to tailor support in online learning settings.

Keywords: Learning Analytics, Machine Learning, Online Learning Environments, Student Engagement, Student Performance Prediction

1. Introduction

1.1 Background of the Study

Over the last twenty years, the nature of education has been subjected to a significant change that could not have been achieved without the major upsurge in the development of online learning opportunities and digital educational services. Quick technological change has also increased this trend, and more lately, global crises like the COVID-19 pandemic, has dramatically transformed the form of content delivery, consumption, and assessment (Dhawan, 2020; Hodges et al., 2020). The authors have estimated that the online-education market in the world would be at least 350

billion dollars in 2025, which is why there is a long-term process of reorganizing the pedagogical paradigms in the current world, which is related to not a short-term crisis demand but rather a fundamental characteristic of the present educational system (The World Economic Forum, 2021). This rapid growth as a consequence has created both opportunities and significant threats both to education facilities all over the world which should not only cope with the challenge of delivering quality course material using digital medium but also maintain the attention of students and their certification of academic success and success.

LMS are the technological basis of modern online education that mediates the course delivery, interaction between learners, and administrative control. Popular solutions emerged, such as Moodle, Google Classroom, Canvas, and others, became common in educational facilities, and overall they communicate with millions of learners and instructors (Turnbull et al., 2021). Every interaction between a learner leaves behind a wealth of digital trace information, including login times and records of resource access through to marking assignments and participating on forums. Although rich, this information is underutilized, however, it can provide significant potential of clarifying and improving student learning experiences (Siemens and Baker, 2012). Knowledge of salient patterns related to the action-taken by a learner, his or her engagement patterns, and future academic results can be identified through the systematic study of such digital footprints.

The introduction of learning analytics as a separate field of study has played a critical role in tapping into the said data reservoir. It can be defined as the organizational measurement, gathering, analysis, and reporting of learning-related data and contextual variables through an intention of explaining and improving pedagogical procedures and learning settings (Siemens and ℕ Long, 2011). This professional field has become the intersection of the study of education along with the data science and computational practices, and has experienced significant traction since its explicit conceptualisation at the First International Conference on Learning Analytics and Knowledge in 2011 (Ferguson, 2012). Learning analytics makes it possible to conduct mass analysis of genuine behavioral data in comparison with conventional techniques of research using mostly self-reporting measures or infrequent tests.

Within the framework of the LMS-based teaching, learning analytics can provide teachers, as well as administration, with a level of insight into the nature of engagement among learners never before. As an example, Moodle has a comprehensive logging system that allows capturing the counts of views of the resources used, quiz answers, and forum posts with a fine-grained time resolution (Romero & Ventura, 2020). Similarly, Google Classroom and Canvas offer learner activity metrics in the form of analytics dashboards which are typically aggregated in aggregate metrics that require further processing to extract actionable pedagogical data. Although these platforms, when combined, allow organizing the system of accruing rich behavioral information that guides the process of instructional design, the formulation of such unprocessed information into evidence-based pedagogical interventions has proved to be a daunting task (Gašević et al., 2015).

1.2 Problem Statement

Despite the rapidly expanding number of online learning options and the amount of data generated within them, one issue exists, namely, the significant percentage of online students facing academic challenges that are not fixed until the end of their study. Empirical studies have

always indicated that the rate of attrition in online learning environments is overwhelming than what would occur in the conventional face-to-face learning setups, with estimates of between twenty and fifty percent depending on the institutional and programmatic factors (Lee and Choi, 2011; Patterson and McFadden, 2009). This can be due to the combination of factors such as feeling lonely, lack of immediate feedback, struggling to control themselves and the lack of a structured support system the physical classrooms are known to have in their habitual nature, which makes it the so-called online learning dropout problem (Hart, 2012).

The implications of the academic struggle in online milieus go well beyond the micro level of student performance, and include the institutional efficacy, rates of program completion, and larger issues of educational equity. Disengagement and the underachievement in students are usually not immediate but proceed with a series of foregoing signs appearing in behavioral patterns away (weeks before academic indicators like midterm examinations) of the shortfalls (Arnold and Pistilli, 2012). However, schools do not often have predictive mechanisms that will be able to detect students at the risk point in time before it is too late to make any meaningful action. Most of the traditional methods used to monitor students are based on periodic assessment and observation by the instructors, which is actually the least effective when using the internet; as an instructor can hardly have the same visibility as to the cognitive process of the learning students (Macfadyen and Dawson, 2010).

The lack of structured early warning systems is a relevant blank in modern educational activity on the Internet. Although individual instructors can identify alarming trends in their own courses, systems with the ability to combine the behavioral indicators in more than one course and identify students with risk-related profiles are few (Arnold and Pistilli, 2012). The gap is notably acute in the context of high-class sizes or where courses are offered asynchronously or where the instruction is provided by the staff that is less experienced in the digital engagement metrics interpretation. Accordingly, analysis and implementation of predictive models used to utilize existing LMS data is a pressing educational need.

1.3 Research Objectives

This work tries to respond to the identified challenges and formulates three major research objectives. The initial aim is to interpolate the student interaction data that it gets when studying on online learning platforms to determine the trends and correlations between engagement behaviours and student results. This involves the orderly mined and processed LMS log data to create valuable behavioural em-pennants to show how students are learning. The second goal is to come up with predictive models which can determine the future academic achievement of students depending on the engagement patterns at early years. This involves using machine learning algorithms on the past data that are made in order to develop models that can detect at-risk students early enough to enable pedagogical action to be taken. The third goal is to identify the learning behaviours that can produce the most significant impact on the academic results, therefore, providing teachers with concrete knowledge about the aspects of engagement that should be given particular attention and stimulation.

1.4 Research Questions

In order to align the investigation with these objectives, the study comes up with three research questions. The initial research question is as follows: What learning behaviors are correlated with

student success of online learning In situations? The question at this stage will be to identify the unique patterns associated with engagement like the frequency of logging in, timeliness in submitting assignments, attending discussion forums, and utilizing learning resources that demonstrate the most positive academic outcomes.

The second research question is the following: How accurately can academic performance be predicted using learning analytics models? This question determines the practical usefulness of predictive modeling by examining the accuracy and the consistency of machine-learning algorithms that predict the outcomes of students upon observing their behavior.

The third research question is as follows: What is the impact of factors on student engagement in online learning environments? This inquiry studies both contextual and individual determinants that inform patterns of engagement by considering that behavior is influenced by an assembly of interacting variables, which comprise motivation, self-regulation skills, and course design aspects.

1.5 Significance of the Study

The relevance of this study can be observed on a number of levels of educational practice and policy. First of all, the timely implementation of the interventions to reduce the percentage of drop-outs by predicting the potential danger among students at the beginning of their education path and providing the necessary treatment promptly is facilitated by the construction of predictive models that help to detect the students at the possible risk and offer the appropriate intervention more quickly. Learning analytics-based early warning would allow spotting students with engagement behaviors that are cause of concern, and grant academic advisors, instructors, and support services the means to intervene prior to academic failure occurring (Sclater et al., 2016).

Second, the paper contributes to the growing trend of using data to make decisions in the education sector through demonstrating how the available LMS data can be utilized to guide the creation of an instruction design and resource distribution and support models of students. Third, salient predictor identification of behavior provides empirical support to refine digital teaching strategies and, as such, allows instructors to design course activities and engagement experiences that align with empirically related behaviors that are linked to student success.

Further on, the given research fills relevant gaps in the body of the literature on learning analytics since it focuses on analysing the application of behavioural data and predictive modelling to the context of genuine online learning. With the investment in the digital learning infrastructure starting to occupy a significant portion of the budgets of the educational establishments, the capacity to extract practical insights out of the information generated by the systems in question, will become an absolute necessity regarding the realization of the potential of online education. In showing this as both possible and useful these predictive analytics in this area, the study provides a basis upon which institutions may consider adopting learning analytics practices, as well as contributing to the overall discussion of how technological tools may be used to support more responsive and student-focused teaching methods in the digital setting.

2. Literature Review

2.1 Learning Analytics in Education

The advent of learning analytics has represented a change of direction in growth trajectory of the fields of educational enquiry and applied pedagogy that fundamentally changes the manner in which the educators understand and react to the process of student learning. Officially, it is considered that learning analytics is the process of measuring, gathering, examining and communicating information about learners and their conditions, to gain insights into and to optimize learning and the learning processes environments (Siemens and Long, 2011, p.��h32). Published in 2011, this definition was the original conceptual framework that informed the future research and development work in this field. The Society of Learning Analytics Research (SoLAR) has additionally stipulated that learning analytics is a specific discursive domain as well as a system of practices that involves the utilization of data to improve learning outcomes, learner achievement, and educational system efficiency (Lang et al., 2017). The main difference between learning analytics and traditional educational research methods is that the former focuses on automated examination of active behavioural data produced in regular learning engagements as opposed to periodic examinations or other self-report scales that might be vulnerable to recall bias or social desirability influences (Gaševic et al., 2015).

The theoretical bases of learning analytics are closely connected with the overall area of educational technology, a discipline that deals with the question of how to foster learning and enhance the performance of students by designing, utilizing and managing the suitable technological procedures and resources (Januszewski and Molenda, 2008, p. 1). Traditionally, educational technology has been focused on the development, designing, and implementing technological tools to facilitate learning; learning analytics expands its role by providing the analytical paradigms guided by which interactions of learners with technical devices and their variations in relation to learning outcomes occur (Gašević et al., 2015). This development is an indication of the maturation of the educational technology practice, where the focus of it is no longer on the questions of access and adoption but on systematic investigation of efficacy and optimization. To the best of my observation, as Ferguson (2012) remarked, learning analytics is the interdependence of various fields which comprise computer science, statistics, education and psychology each has a contribution of methodology and theoretical orientation which add more insights on the process of digital learning.

One of the key areas of discussion in the research of learning analytics is the data gathering and processing created by Learning Management Systems (LMS). Such applications, as Moodle, Blackboard Learn, and Canvas, have become the key sources of data to research learning analytics studies as they have all the features of logging and are implemented in educational institutions worldwide. LMS systems provide granular data of just about every student interaction with pageviews, resource visits, quizzes attempts, assignment submissions, and forum posts, each interaction is normally recorded with time and user ID (Romero and Ventura, 2020). This information, commonly referred to as digital trace data, provides scholars with a level of insight into the student learning mentalities that they never had access to before using traditional classroom observations or self-reporting structures. Nevertheless, LMS platforms produce log data that is not usually structured or in small amounts, and processing it to render it into analytic variables is generally intensive. Researchers have established different models of gathering and combating LMS data into student engagement measurements, such as frequency, duration,

frequency of recency, intensity, and frequency of patterns of interaction (Baker and Inventado, 2014).

Student engagement analytics forms one of the most current fields of interest among learning analytics literature because of the long-established correlation between engagement and academic performance. The concept of engagement, which can be described as the level and amount of student participation in the learning process, has long been identified as a highly important indicator of academic performance in the traditional learning setting as well as online learning (Kuh, 2009). In online learning, engagement is observed in terms of behavioural patterns which are measurable and can be profiled using LMS data. Henrie et al. (2015) have suggested a multidimensional approach to the measurement of engagement via learning analytics, selecting three different dimensions: behavioural engagement (e.g., the frequency of logins, the frequency of accessing resources, the frequency of assigning them), cognitive engagement (e.g., the duration devoted to a learning task, the intensity of interaction with resources, metacognition behaviours), and emotional engagement (e.g., expressed sentiment in the posts in the forums, affective reactions to feedback). The ability to continuously quantify these dimensions using automated analysis is a big improvement over traditional engagement measurement approaches which, in most cases, are based on periodic surveys, teacher observations, or self-reports that are often done infrequently and might have a narrow scope (Azevedo, 2015).

Educational data mining (EDM) represents a closely related study that has a lot of overlap in terms of methods and concepts with learning analytics. Whereas learning analytics is more inclined towards the utilization of the analytic results in the educational practice and decision making, educational data mining is more interested in the creation and implementation of the computational algorithms of theuncovering of patterns in educational data (Baker and Inventado, 2014). Some of the common methods in EDM are the use of clustering to identify student subgroups that share a similar pattern of behaviour, classification to predict categorical measures such as course completion, relationship mining to establish a relationship between the variables and text mining to interpret student writing and forum posts (Romero and Ventura, 2020). The interaction linking learning analytics with EDM has been thought of as complimentary, as opposed to competitive, with EDM creating methodological rigour and computational methods whilst learning analytics creates educational context, theoretical basis, and pragmatic orientation in intervention and enhancement (Siemens and Baker, 2012). Collectively, these disciplines have been part of an expanding literature on the topic of the potential to educators to use digital learning traces to inform educational practice and help students achieve success.

2.2 Predictive Models in Education

Creation and use of predictive models of student performance are one of the key methodological issues in learning analytics, which makes it possible to identify learners at risk of academic trouble in a timely manner and provide interventions that can be implemented in time. Here, the goal of predictive models is to predict academic performance academic final course grades, course completion status or dropout risk using early behavioural data, which are gained out of learning management system (LMS) data (Joksimović et al., 2016). An extensive body of literature has examined the relative performance of different modelling methods, and important

trade-offs between predictive accuracy, interpretability, and applicability have been found, which have to be trade-offs in choosing the analytical methods.

Logistic regression has continued to be one of the most used predictive modelling tools in learning analytics due to its reflective merit and solid theoretical underpinnings in statistical inferences. The logistic regression can be described as a generalized linear model that is appropriate when binary classification is necessary: it assumes the probability of a dichotomous outcome e.g. passing a course versus failing a course is a linear combination of independent variables (Hosmer et al, 2013). The estimates drawn out of logistic regression can be interpreted and give an approximation of the direction and the strength of associations between behavioural indicators with academic outcomes and it is therefore the most useful method to determine which engagement behaviours have the most significant impact on student achievement. The usage of logistic regression in learning analytics studies always proves to be effective, and predictive accuracy is correspondingly high and measures between 0.70 and 0.85 on the area-under-the-curve (AUC) scale (Tempelaar et al., 2015). However, logistic regression assumes linear association among predictors and the log-odds of the outcome, which might limit its ability to detect complex and non-linear patterns of interaction that could have taken place in educational data (James et al., 2021).

Second, decision tree algorithms also provide a more intuitive methodology with related predictive power and visualisation of the decision rules to make them easier to access to educators and practitioners. Decision tree works by recursively dividing the data into smaller and smaller homogenous groups according to predictor variables hence forming a tree-like structure where each internal node is actually a decision rule that used on a certain variable and each leaf node is a predicted outcome (Quinlan, 1986). The resulting models are quite easy to interpret and the stakeholders can know the definite patterns of behaviour that predict certain outcomes. An example provides that a decision tree will indicate that the students who fail to complete their assignments with a third week of a semester and those who have less than three times a week as average frequency of logging in are highly at risk of failing in the course. Yu and Jo (2014) have shown that decision tree models are effective in the prediction of performance among the students in the online learning settings, with a high accuracy rate of over 80 percent. Nevertheless, decision trees can be overfitting, which may occur especially when they are trained in small datasets or when there are many predictor variables in comparison to the sample size; thus may restrict their ability to generalise to new data (James et al., 2021).

The advantage of random forest algorithms over single decision trees is that they build an ensemble of multiple trees, which are trained on bootstrapped datasets and aggregate their answers either by voting or averaging. The idea is proposed by Breiman (2001), and it minimises variance and increases the predictive stability in two manners. The first is the bootstrapping to more varied training samples and the second is the random selection of features at each split to decorrelate single trees. This is because random forest can be used to align the most complicated non-linear relationships, manage high-dimensional data with a large number of predictors, as well as provide estimates of variable importance, helping to determine which engagement captors provide the most significant contribution to prediction (Hellas et al., 2018). Research in the learning analytics setting commonly has findings, not to mention comparative ones, that random forest attains higher accuracy compared to logistic regression and single decision trees (Conijn et al., 2017; Hellas et al., 2018). Another definite successful feature of random forest is

that it is considerably more resistant to overfitting in comparison with single decision trees with the price of much less interpretability of the obtained models, a compromise between predictive power and the explanation.

The other influential machine learning method used in performance prediction tasks of students is the Support Vector Machines (SVMs). SVMs are worked out based on mapping of the data to a high-dimensional feature space and identifying the best hyperplane that optimally divides distinct classes as well as increases the margin between the hyperplane and the nearest data points of each of the classes (Cortes and Vapnik, 1995). Using kernel functions, SVMs are capable of dealing with non-linearly separable data due to implicit mapping of the input features into higher dimensional space in which linear separation becomes possible. A study by Osmanbegovic and Suljic (2012) revealed applicability of SVM in predicting student performance with accuracy rates that are at par or higher than other machine learning algorithms in various learning data set. But in most cases, SVMs have low interpretability compared to logistic regression or decision trees, and their performance depends on parameter choices, such as kernel function and regularisation, which means that carefully tuning them is necessary to obtain the best performance (James et al., 2021).

Implementing such predictive models is enabled by the use of strong programming libraries and programming tools, which are now common in the learning analytics research. The R programming language has become a popular platform to conduct statistical computing and graphics in educational research and provides wide packages of implementing predictive model using libraries as glm, logistic regression, rpart, decision trees, randomForest, random forest models, and e1071, support-vector machines (R core team, 2021). The comprehensive data manipulation, data visualisation and data reporting ecosystem of R makes it especially suitable to iterative and exploratory characteristics of learning analytics research. Likewise, Python has become prominent in the field of learning analytics, where there are libraries such as scikit-learn, which has many machine-learning functions, pandas used to process data, and numpy to do numerical calculations (Pedregosa et al., 2011). The availability of these tools has greatly reduced the technical issues involved in predictive modelling and thus educational researchers are able to use more advanced methodology of analysis without the need of specialised training in computer science.

2.3 Online Learning Environments

Online learning settings are inherently defined by the properties of the data to be analyzed and the trends regarding the interest of students. Learning Management Systems like Moodle and Blackboard Learn are the main platforms on which online education is provided, thus forming the main source of data to be used in learning analytics research. Because Moodle is open-source software, it has played an exceptionally persuasive role in learning analytics research since it has transparent architecture and extensive logging and has an engaged development community that continuously improves its analytics functionalities (Romero and Ventura, 2020). The platform captures rich event logs of every interaction among the users, which helps the researchers to salvage the student learning paths with a fair amount of temporal and categorical depth. Commercial poster alternatives, such as Blackboard Learn, which is extensively used in North American higher education, have similar functionality in logging and this can be studied by many learning analytics (Coates et al., 2005). The two, along with others including Canvas offer

application programming interfaces (APIs) that enable the researcher to access log data to analyze it, thus leading to the creation of custom analytics programs and also integrating third-party analytical packages.

The asynchronous learning is also a characteristic attribute of numerous internet based learning platforms, which makes them distinguishable in comparison with the conventional face-to-face classroom environment and the option of a synchronous internet-based course. Unlike in synchronous learning, which occurs in real time with both students and instructors present, asynchronous learning allows students to study course materials, and undertake learning activities more at the flexible time and space intervals (Murphy and Rodriguez-Manzareas, 2009). The temporal flexibility is generally recognized as the key benefit of online education that allows the student to balance academic activity with job, family, and other responsibilities and fits various phenomenon of time learning. However, asynchronous quality of most web-based classes poses a great challenge to student participation and continuation. A lack of scheduled classes meetings can reduce the feeling of accountability to the point of team members creating undisciplined procrastination and such immediate feedback may not be provided meaning that the learners might not know their progress and comprehension (Hart, 2012). These issues make learning analytics especially useful in the asynchronous environment, where the systematic analysis of the traces of digital engagement can provide educators with an insight into the previously invisible processes of learning.

Student interaction logs are the raw data that participates in generating engagement indicators in the learning analytics research. Each activity that a learner does inside LMS is logged with time, resource name, type, and usually, one or more additional metadata details (duration of interaction or browser details, etc.) (Macfadyen and Dawson, 2010). An example of the events that are logged might include seeing a course page, watching a lecture video, writing an assignment, leaving a post in a discussion forum, or taking a quiz. The time aspect of such logs is rather helpful because it allows the researcher to study the patterns of engagement across time, such as frequency, regularity, time in reference to deadlines, and distribution of activity across the course time (You, 2016). A study conducted by You (2016) established that temporal dynamics of LMS usage, especially frequency of the timing of logging in and time between study sessions, is a predictable academic success even after considering the mere volume of activity. Such results suggest that knowledge of the mode of engagement of learners, both in terms of pattern, rhythm, and time allocation of your activity, can be as important to comprehend as the extent of your engagement.

2.4 Key Variables Affecting Student Performance

Numerous empirical studies have been conducted to investigate the correlation between behavioral variables inferred with the help of learning management system (LMS) data and academic performance among the students, identifying a number of engagement predictors that are consistently able to accurately predict outcomes. Of these, the frequency of logging in has proven to be one of the strongest foretellers of success when online computing. Several articles present significant positive associations between the number of LMS visits and the end course scores (Macfadyen and Dawson, 2010; Morris and others, 2005). The frequency of logging in is a proxy of general interaction and continuity, which includes abundance of investment students make in the activity of the course and retention to the digital environment. However, scholars

worry that easy frequency indicators will not capture the quality of engagement because they cannot distinguish between a meaningful interaction with course material and a superficial engagement with the content that only includes frequent updates taking every time (Gašević et al., 2016). Less crude measures of student engagement, like the frequency of logging in measured through entropy or any standard deviation of the interval between logins, and the pattern of access to a wide array of course materials, might provide more information about the qualitative aspects of engagement.

The time of submission of assignments and especially the submissions in relation to the deadline has become another important predictor of online learning student performance. According to Park and Jo (2015), students who submit their assignments early enough before the deadline have better grades in comparison with those who submit at the last minute or late as it is. Such a relationship, probably, is indicative of differences in time-management ability, conscientiousness, or preparatory activities in generalisable learning environments. Further, the time distribution of submissions of assignments during a course can be used as a predictive variable; when the submissions of assignments decrease or the submissions are becoming later and later, this is an early indication of disengagement or academic and learning challenge. Research has shown that consideration of assignment submission pattern whether students are consistent or progressive in terms of timing of submission is a useful information that can be relied upon in predicting the end results (You, 2016).

The participation in discussion-forum has been widely studied as indicators of engagement and predictive of more academic achievements in an online course. It has been linked to enhanced learning, increased communicative feeling and better grades as a result of active contribution in forums (Dawson, 2006). Nonetheless, the interaction between the activity in forums and the performance is more complicated, and there are indicators that the number and quality of engagement play a crucial role. Studies have shown that not every post made in the forums is equally predictive; posts that require higher-level thinking, active interactions with course material and dialogue with fellow students positively correlate with learning outcomes as compared to agreement posts and other logistical questions (Wise et al., 2012). Also, the timing of engagement in the course up to the point of the course progress affects its predictive validity, early engagement in the discussion relates to positive performance in comparison to engagement at the very end of the course or after receiving the performance feedback (Joksimović et al., 2016).

Viewing behaviour of video-lecture has been very popular when video based learning is on the rise in greater magnitude online. It has been demonstrated that student performance is predictable by the number and trend of the number of videos watched (Kim et al., 2014). Students who get more of lecture information, who do not skip parts of the videos but wait till the end but do not switch patterns of watching videos are more likely to get better grades. The more sophisticated interaction video analytics, such as pause events, replay events, navigation, and the speed of viewing, can further inform the understanding of videos as an instructional strategy and what metacognitive activities are linked to effective learning (Giannakos et al., 2015). These sophisticated video interaction measures are the distinguishing tool between active service of the instructional material and passive watching of the video.

The time spent on learning resources is one of the primary indicators of how engaged a student is, yet its dependence on academic success is complex and should be interpreted with attention. The overall LMS time is positively correlated with grades, however, studies indicate that both the quality and allocation of time are more decisive than the amount (Kovanovic et al., 2015). Learners who put in an equivalent study time per course are more likely to perform better than their counterparts who study in great quantities and at the last moment due to the deadline of an assignment. Additionally, time and performance relationship might be curvilinear: moderate levels of time spent indicate efficiency in learning, but excess time might show that the person struggles or is inefficient. Kovanović et al. (2015) demonstrated time on task measures to have greater predictive validity in the context of the nature of learning activities under way so that time on higher-order learning tasks like problem-solving and reflection have stronger relationships with learning outcomes in comparison to time spent on the passive consumption of contents.

2.5 Research Gap

Despite the fact that significant amount of literature research has been conducted to study the concept of learning analytics and predictive modelling in the educational sector, a number of gaps still exist to challenge the extrapolation of the existing findings and their application. The clear weakness of the study is a geographical bias of the existing literature, where most learning-analytics studies have originated in developed economies, specifically in North America, Western Europe, and Australia, and limited research has been conducted on online learning settings located in the developing world (Viberg et al., 2018). Such spatial localisation, on the one hand, compromises the generalisability of the recorded results since the heterogeneity of the educational systems, technological support, student population, cultural tropes, and socioeconomic settings in the differences in territories is quite pronounced. The teaching challenges faced by institutions in developing nations such as limited technological access, intermittent internet, size of the enrolment group, and the presence of heterogenous student bodies with diverse digital abilities may require more analytical modalities, more specific model forms, and different intervention plans than those developed in more prosperous settings (Tlili et al., 2020).

The second gap relates to the development of heterogeneous streams of behavioural data in sound predictive paradigms. Although several studies have revealed the predictive power of single behavioural metrics, e.g. frequency of login and timing associated with assigning a work, forum postings, video-viewing, and so on, only a small number have explicitly conducted a study enabling the combination of these variables that captures temporal, interactional and multidimensional complexity of learning (Joksimović et al., 2016). Joksimovic et al (2016) argued that most of existing learning-analytics models are based on aggregate and static measures of behaviour and are not able to capture the dynamics of engagement changes over time and inter-relations between various dimensions of engagement. In line with this, the design of models that combine temporal dynamics and explain changes in engagement processes and summarize interactions among different behavioural signals continues to be a promising field of research.

Third, the combination of behavioural data with predictive modelling under non-Western educational conditions is understudied. Though the literature reviewed has examined individual

measure of behaviour separately, there has been minimal systematic integration of varied behavioural measures on large scale predictive models which are specifically oriented to designing in the context of developing countries. An approach to the methodology that combines LMS authentication logs with other categories of data streams - including institutional data, demographic polls, and learning-resource analytics - could provide a more holistic view of the factors of student success in those settings. The given investigation is thus aimed at closing these identified gaps by developing predictive models that consistently integrate various behavioural indicators inside the frame of an online education system that serves in a developing country and thus contributing to the learning-analytics corpus in general and solving context-specific pedagogical issues.

3. Conceptual Framework

A conceptual framework serves as the frame of empirical investigation that specifies the main constructs that are being studied and how they may be anticipated to relate to one another (Miles et al., 2020). The framework, in the current research, describes the connections, in which four behavioral engagement variables, that is, the frequency of logs, assignment submission rate, forum engagement, and the use of learning resources, are independent predictors, and the outcome variable, student academic performance, quantified by grades or grade point average (GPA). The framework is premised on the self-regulated learning theory, which offers a theoretical take to explain how proactive control over their learning behaviors by students can be converted into academic success at the online learning settings.

3.1 Theoretical Underpinning: Self-Regulated Learning Theory

The Self-regulated learning (SRL) theory has a strong explanatory base to the relationship expectations. SRL as a theoretical construct derives its basis via the seminal contribution of Bandura (1986), Zimmerman, (2002), and Pintrich, (2004) as it argues that effective learners actively regulate cognitive, motivational as well as behavioral processes to achieve individual learning objectives. It can be observed that SRL in digital environment will be represented by the behavioral patterns of planning its study sessions, tracking its understanding, consulting services on demand and adaptability to its strategies involving the feedback (Winne and Hadwin, 2008). These processes are not completely internal, but on top of it create digital footprints which can be collected using the Learning Management Systems (LMS). Consequently, the behavioural variables that have been selected for this inquiry (login frequency, assignment submission rate, forum participation, and learning resources utilisation) are conceptualised on the basis of observable indicators of underlying self-regulatory processes.

Self-regulating learning theory emphasises the idea that learners that set objectives, monitor progress and strategically engage with resources, are more likely to achieve favourable academic outcomes (Zimmerman, 2002). In online learning environments, where the external scaffolding tends to dampen as compared to a face-to-face learning environment, the role of self-regulation takes on an increased importance (Broadbent & Poon, 2015). Students who have high access to the LMS show evidence of goal orientation, persistence, those who submit their assignments on time show evidence of effective time management, participation in forums reflects help-seeking and social, while active utilisation of learning resources reflects cognitive engagement with the

content. Each of these behaviours is related to the cyclical phases of self- regulation: forethought (planning), performance (volitional control), and self-reflection (Zimmerman, 2002).

3.2 Independent Variables

The frequency of logins can be described as the measure of how many times a student logs into LMS within a given time frame, which is usually either daily or weekly or the course duration. It has always been evident that regular and frequent logins have a positive correlation with academic success in online learning (Macfadyen and Dawson, 2010; You, 2016). In SRL terms, the frequency of logging in can be explained by the fact that forethought and performance stages: students who log in regularly have higher probability to stay updated with the course requirements, plan their time of studying, and ensure constant engagement thus avoiding accumulation of backlog and cognitive overload (Broadbent and Poon, 2015). Further still, the regularity of the login, not sporadic bursts, is a demonstration of the stable self-regulation and proved to be a better predictor over a total number of logins (Gašević et al., 2016).

The assignment submission rate is a metric that describes the lateness and the level at which students are able to complete the necessary assessments. The variables are operationallyized by the ratio of assignments handed in on time or the daily lag (in days) of the deadline. A direct application of performance-control stage of self-regulation especially time management and goal adherence is assignment submission behaviour (Zimmerman, 2002). Students that make sure to hand assignments on time effectively plan their work and can develop resistance towards procrastination. According to a study conducted by Park and Jo (2015), one of the strongest predictors of final grades was the submission time and the earlier or on-time submission was associated with better grades. On the other hand, submitted late and missed are the common indicators of self-regulation failure and the antecedents of academic danger (Arnold and Pistilli, 2012).

Forum participation is used to describe the student contribution to discussion forums in the LMS. It may be quantified using the quantity of posts, responses or topics started and more by the quality of the contributions (Wise et al., 2012). In self-regulated learning theory, help-seeking strategies and social interaction are related to the forum participation which is deemed as an essential self-regulatory behaviours (Pintrich, 2004). Good, self-regulated learners actively seek elucidation, exchange information, as well as participate in the co-creation of knowledge. Such behaviour is recorded on online forums. Research has demonstrated that students who are active in discussion rooms have better grades and record less dropout rates (Dawson, 2006; Joksimović et al., 2016). Furthermore, the aspects of participation by time (early or late in the course) shed some light on the proactive and reactive involvement.

Resource usage LRU covers the level and format with which the students engage with course content including lecture videos, readings, quizzes, and other types of content. These might be the amount of time spent, the amount of resources visited, the strength of engagement (e.g., video completion rate, rereading material), and the spread of access among types of resources (Kovanovic et al., 2015). This dimension of self-regulation is the cognitive aspect, which is captured by this variable. Winne and Hadwin (2008) remark that adaptive users of resources who strategically monitor their comprehension of resources are more likely to achieve deep learning. The study by Kim et al. (2014) and Giannakos et al. (2015) did indeed prove that the greater the resource usage, especially at the same time intervals, the greater the academic outcomes. Within

the framework of SRL, the monitoring and adaptation processes are reflected in how a student uses resources: metacognitive awareness is manifested in cases where the student is ready to reconsider the challenging content or employ interactive resources.

3.3 Dependent Variable

The performance of students academically is measured by the final course grade or cumulative GPA the student acquires. Finally grades are the most common summative measure of achievement in web-based learning contexts and are the canonical outcome gauge in a research of predictive modelling studies (Romero and Ventura, 2020). It is justified to use the final grades as the dependent variable due to its immediate relevance to institutional accountability and student progression and practical goal of identifying at-risk learners early as possible. Moreover, grades reflect the cumulative effect of self-regulated behaviors of students throughout the duration of the course, which makes them an appropriate measure of model validation (Tempelaar et al., 2015).

3.4 Relationships Among Variables

The hypothesized conceptual model assumes a positive direct correlation between all the independent variables and the academic performance of the students. To be more specific, it is hypothesized that an increased rate of finishing the logins (i.e. submitting tasks promptly), higher rate of rate of assignments submission (i.e. submitting them on time), more forum activity, and heavier use of learning resources is all correlated with high end grades. Such interrelations are based on the self-regulated learning theory that indicates that the relations are not illustrious and are manifested as a coherent pattern of student agency and metacognitive control (Zimmerman, 2002). Moreover, the framework recognizes that the independent variables can have a relationship; that is, when students log in regularly, they have more chances of accessing the learning resources and using forums. Nevertheless, the main point of analysis is the aggregate predictive value of the variables since machine-learning models will use all the variables at once to estimate the overall performance.

It also gives the possibility of nonlinear relationships and interactions to be provided through the framework. An example is that the impact of frequency of logins can be neutralised at a tipping point, or a combination of infrequent forum use and tardiness could be particularly a marker of danger. Such complexities will be addressed by the use of ensemble techniques like random forests that can generate a non line of interaction without on need to specify the interaction (Breiman, 2001).

3.5 Visual Representation

Although an illustrative representation is not possible within the context of this textual presentation, the given conceptual framework can be presented in the form of a guided diagram that includes four antecedent variables: Login Frequency, Assignment Submission Rate, Forum Participation, and Learning Resource Usage and all relate to a single outcome variable, which is Student Academic Performance. Summarizing this schematic, one of the overlay on these schematics is the result of self-regulated learning theory which states that observable behavioral indicators are the products of underlying self-regulatory processes i.e. goal setting, strategic planning, monitoring, and adaptation. Moreover, the schematic includes directional connections

between the antecedent variables to consider its common use in the empirical relationships of variables.

3.6 Justification of Variable Selection

Their choice of the four independent variables is based on the theoretical factor and empirical facts regarding learning analytics research. Theoretically, the combination of the three variables help put into perspective the three main stages of self-regulated learning: forethought (frequency of login and planning), performance control (assignment submission and resource use) and self-reflection (forum participation as a form of feedback-seeking) (Pintrich, 2004). They are empirically the most commonly found predictors in studies that use LMS and have shown stable results in their relationship with academic outcomes across different situations (Macfadyen and Dawson, 2010; Park and Jo, 2015; You, 2016). Additionally, these variables can easily be found in the standard LMS log data, which makes the framework feasible to apply in the education settings in the real world. With these core behavioural indicators, the study will establish accurate and interpretable predictive models that will help in the development of the earliest warning systems that will lead to the provision of time-sensitive pedagogical intervention.

Overall, the given conceptual framework shapes the given research on the basis of self-regulated learning theory and provides a clear outline of the hypothesized connections between the observable LMS behaviours and the student academic performance. It gives a systematic framework in how data has to be gathered, analyzed and interpreted so that the predictive modelling attempts portfolio is governed by an organized theoretical point of view.

4. Methodology

4.1 Research Design

The current study follows the quantitative research design in which learning analytics research processes are combined with predictive modelling in order to explore the connection between behavioural engagement and academic performance of students who are studying in virtual learning environments. The quantitative research designs are characterized by the systematic gathering and analysis of the numerical data in order to recognize the patterns, test hypotheses, and extrapolate the results to the larger populations (Creswell and Creswell, 2018). The choice of a quantitative methodology, in its turn, is especially relevant to the research questions, which are designed to determine predictive relationships between the measurable behavioural predictors and the quantifiable academic performance, as well as to demonstrate the level of prediction accuracy through statistical measures of predictive models.

Two analytical approaches which mutually support one another are used in the research design; data mining and statistical analysis. Data mining is defined as the qualitative procedure of extracting trends, associations, and outliers in large data volumes, in most cases with the aim of making actionable consequences (Baker and Inventado, 2014). Here, the data mining techniques will be used on raw learning management system (LMS) log data to condense relevant behavioural indicators, identify engagement patterns and develop predictive models using machine-learned algorithms. Data mining is enhanced by statistical analysis providing inferential processes to test assumptions according to the association between variables and affirm the

statistical significance of the noticed patterns (Field, 2018). Together, these approaches enable an in-depth investigation into the extent to which the engagement behaviours predict academic performance to achieve the goals of the exploratory type examined by discovering patterns and the confirmatory type examined by proving a hypothesis.

The design is clearly predictive, as opposed to purely descriptive, and the goal is to achieve the model in such a way that it can predict outcomes of students, depending on the early indicators of engagement. Predictive research designs are currently more common in the field of learning analytics, as they allow discovering at-risk students in a timely manner to then make informed choices about the type of intervention that should be implemented (Siemens and Baker, 2012). The practice will expose the useful usefulness of learning analytics in educational practice by using past data on a completed group to the training and validation of predictive models that, in turn, can be applied to new groups to be adequate proof of the research concept.

4.2 Data Source

The main source of information that will be used to carry out research in this study is a Learning Management System (LMS) that will be used in a university setting. The research will utilize statistics obtained through the use of Moodle, which is an open-source LMS platform, and has become popular at numerous institutions of higher learning throughout the world. Learning analytics research can be achieved using Moodle especially due to its mature logging architecture, which accumulates exhaustive event logs related to each user interaction, and a proactive development community that supports analytics-related extensions (Romero & Ventura, 2020). Even though most of the alternative platforms like Canvas LMS can be termed as dependent on the availability of these in the relevant institution, Moodle is chosen because of the extensive logging options it provides and its wide adoption.

The information that will be harvested in the LMS will be in the form of three major sets of data. First, logs of student activity will provide the description of all interactions of students in the course with time, type of action and identifiers of resources. The basis of computing the indicators of behavioral engagement is based on these logs, including frequency of logins, pattern of resource access and participation in forums. Second, assessment scores will involve the grading of all graded tasks given during the course, which will involve quizzes, assignments, examinations, and other summative assessment. The scores of assessment will be used as the predictors of final outcomes and the dependent variable in the validation of the model. Third, the data about course completion will provide the information about the final results, such as the data about the successful course completion and the final grades of the students.

The sample of the study will consist of about 200 to 500 university students in one semester of a course that is offered in entirely online or in a hybrid mode. This is a sufficient sample size, as other learning analytics researchers have indeed been able to build predictive models with sufficient statistical power (Conijn et al., 2017; Macfadyen and Dawson, 2010). By choosing one course to analyze, one opens up the variables at the course level: the design of the instruction, the structure of the assessment, teacher influences, etc. that would otherwise confound the correlations between behavioral measures and performance. The data collection will be conducted over a period of at least one academic semester (the average number of weeks is 14-16), to have the full picture of student engagement throughout the course period, beginning at the moment of its start and continuing to the final review.

4.3 Data Collection

Systematic retrieval of logs of interactions and assessment records of students on the LMS database was used in the data collection process. All the extraction was performed in regard to institutional ethical considerations and data protection laws, whereas data were anonymized before the analysis in order to preserve privacy of students.

Student interaction logs constitute the largest and most fine-grained type of data that consists of logs of all activities undertaken by individual students in the LMS.

The common elements of a log include student identifier, time by which they took the action, type of activity (e.g., course view, resource view, quiz attempt, forum post), resource identifier and some contextual data of the learning activity (Romero & Ventura, 2020).

These were logs that were obtained at the end of the semester and therefore they covered the whole course period. The results of the assessments were read out of the LMS gradebook, and there is the record of all the graded activities. In the case of every student, data consisted of scores on individual assignments, quizzes, exams, and other graded elements and the final course grade.

Interactive logs were related to the assessments data by means of anonymity-based student identifiers which allowed combining both behavioural and performance data. The interaction logs, which, in turn, are based on the raw Log data, produced engagement indicators through the use of feature engineering, which summarizes raw log information into substantial behavioural metrics. In particular, speaking of the individual students, the following engagement measures were calculated: frequency of logins (the total number of logins, the consistency of colleagues, the nature of the disciplines), assignment submission rates (dates of submission in relation to a deadline), forum use (amount of posts, responses, and threads made), and use of learning resources (the total amount of time spent, number of different resources used, and intensity of interactions).

All the data were stored in safe database with relevant access restriction features. To record the definitions of the variables, coding schemes and any transformations that were carried out during preprocessing, a data dictionary was kept. Data cleaning was also used in the data collection stage to overcome any missing data, outliers and discrepancies, which might jeopardize the analytical validity.

4.4 Data Analysis Techniques

The data analysis stage will use a combination of statistical approaches and machine learning techniques in order to answer the research questions. The analysis will run in a progressive way starting with descriptive statistics, correlation analysis, and multiple regression and ends up on developing and evaluating machine learning predictive models.

Summary statistics will be calculated to describe sample characteristics and the distributions of key variables. Descriptive statistics (mean, median) and measures of dispersion (standard deviation, range) will be computed for all continuous variables including login frequency, assignment submission rates [12], forum participation counts [13], resource usage metrics [14]

and final grades. Categorical Variables: Frequency distributions will be analyzed. Descriptive analysis is also important to describe the sample, detect outliers and check whether distributional assumptions are in accordance with future inferential analyses (Field, 2018).

Bivariate correlation analysis will be used to analyze the relationships between each of the behavioral indicators and academic performance, and within the behavioral indicators. Pearson product moment correlation coefficients will be calculated for normally distributed continuous variables and Spearman's rank correlation for those variables that do not meet the assumption of normality. Correlation analysis gives first insights into the engagement behaviors most strongly correlated with student outcome measures and serves as a diagnostic function to identify potential multicollinearity between predictor variables, that will bias future regression or machine learning models (Tabachnick & Fidell, 2019).

A multiple regression analysis will be used to assess the joint and independent contributions of academic performance predictors while controlling changing potential confounding variables. Multiple regression is thus appropriate for modeling the relationship between multiple independent variables and a continuous dependent variable (final grade), yielding interpretable coefficients that indicate the direction and magnitude of each predictor's effect (Hair et al., 2019). We will specify the regression model as: Final Grade = $\beta 0 + \beta 1$(Frequency of Logins) + $\beta 2$(Rate of Assignment Submissions) + $\beta 3$(Participation in Forum) + $\beta 4$(Usefulness of Learning Resources) + ε. Standardized coefficients (beta weights) will be reported to allow for comparisons of relative importance of the predictors.

According to predictive accuracy, machine learning prediction models will be developed for the research question. We will implement and compare four algorithms: Logistic Regression (for passing or failing the selection criteria), Decision Trees, Random Forest, and Support Vector Machines. These algorithms were chosen for their proven performance in previous learning analytics studies and ability to model various types of relationships within the data (Conijn et al., 2017; Hellas et al., 2018). The analysis will be executed in Python (using the scikit-learn library) and R (caret package), which are both commonly used learning analytics research tools (Pedregosa et al., 2011; R Core Team, 2021). Descriptive statistics, correlation analysis, and multiple regression will be conducted using SPSS.

Stratified random sampling will be used to split your dataset into training (70%) and testing (30%) sets, keeping the distribution of outcomes consistent across both splits. The models will be trained on the training set and their predictions used to evaluate themselves with the held-out testing set, which provides estimates of predictive performance that are unbiased. For each algorithm hyperparameter tuning will be applied to optimize for model performance using cross-validation on the trainset.

Model evaluation will employ several metrics appropriate for classification tasks, as summarized in Table 1.

Table 1: *Evaluation Metrics for Predictive Model Performance*

Metric	Definition	Formula	Interpretation
Accuracy	Proportion of correct predictions	(TP + TN) / (TP + TN + FP + FN)	Overall correctness of the model; suitable when class distribution is balanced

Precision	Proportion of positive predictions that are correct	TP / (TP + FP)	Measures exactness; high precision indicates low false positive rate
Recall (Sensitivity)	Proportion of actual positives correctly identified	TP / (TP + FN)	Measures completeness; high recall indicates low false negative rate
F1-Score	Harmonic mean of precision and recall	2 × (Precision × Recall) / (Precision + Recall)	Balanced measure for imbalanced classes; penalizes extreme values

Note. TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives. Evaluation metrics derived from Powers (2020).

Then, these measures will be calculated for each predictive model on the testing set allowing a structured evaluation when comparing algorithm performance. Thereafter for Random Forest models, measures of variable importance will be extracted to investigate which behavioural indicators provide the largest contribution to predictive accuracy, directly addressing the research question on key predictors of performance in students.

All analyses will be performed while considering the assumptions of each statistical technique. For regression models normality, homoscedasticity and independence of errors will be checked through residual diagnostics. Cross-validation results will be inspected for stability and generalisability in the machine learning models. By blending statistical and machine learning methods, we combine the interpretability of traditional statistical approaches with the predictive strength of sophisticated machine learning algorithms in a unified analytical frame.

5. Results and Findings

In this part of the paper results are given based on analysis performed with data collected from interactions of students in Moodle learning management system. After data cleaning and removal of incomplete records, the final dataset included 384 undergraduate students enrolled in a semester-long online course. The analysis followed three steps: descriptive statistics and correlation analysis to characterize engagement trajectories; building and evaluating predictive models; identifying key behavioral predictors, computation of engagement trajectories, and graphical representation. Statistical analysis All statistical analyses were performed by means of Python (scikit-learn, pandas) and R (ggplot2, caret), using $\alpha = 0.05$ for significance testing.

5.1 Model Prediction Accuracy

Four machine learning algorithms were trained and validated to predict student academic performance operationalized as a binary outcome (pass or fail) based on a final grade cut-off threshold set at 60% Logistic Regression (LR), Decision Tree (DT), Random Forest (RF.) and Support Vector Machine (SVM). The data was randomly subset into a training set (70%, n = 269) and testing set (30%, n = 115), retaining the class distribution (overall pass rate 78%). Hyperparameters were adjusted with 10-fold cross-validation on the training set, while performance assessment was calculated over the held-out testing set using accuracy, precision, recall and F1-score.

Table 2: *Predictive Performance of Machine Learning Models on Testing Set (n = 115)*

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.826	0.85	0.93	0.89
Decision Tree	0.809	0.83	0.92	0.87
Random Forest	0.870	0.90	0.94	0.92
Support Vector Machine	0.843	0.87	0.93	0.90

Note. Precision, recall, and F1-score are reported for the "pass" class. Metrics calculated using scikit-learn (Pedregosa et al., 2011).

Random Forest outperformed other algorithms with the highest overall accuracy (87.0%) and F1-score (0.92). Indeed, its best performance is in line with earlier work that found ensemble methods successfully capture non-linear interactions between engagement variables (Conijn et al., 2017; Hellas et al., 2018). Logistic Regression, though a bit less accurate (82.6%), served as an excellent baseline and was also more interpretable. Decision Tree had good recall (0.92) but low precision (0.83), suggesting that it tends to over-predict the pass class. Although all SVM metrics were good, SVM needs a very large hyper parameter optimization for the model stability.

The high recall values (0.92–0.94) across all models suggest that the behavioral indicators as a whole are sensitive in identifying students who eventually pass the course. More importantly, false negative rate (students predicted to fail who actually passed) was low (6–8%), indicating that an early warning system based on these models would rarely fail to detect at-risk students. But precision was a bit worse for Decision Tree (0.83), which means that this model would trigger a higher number of false alarms, resulting in potentially unnecessary interventions (higher costs in term of health quality).

5.2 Key Predictors of Student Performance

To determine which learning behaviors had the greatest impact on academic outcomes, we extracted variable importance measures from the Random Forest model that internally ranks predictors based on mean decrease in impurity (Breiman, 2001). Standardized coefficients from the multiple regression model (which used final grade as continuous outcome variable) were also assessed to verify directionality and magnitude of effects.

Table 3: *Variable Importance Rankings from Random Forest and Standardized Coefficients from Multiple Regression*

Predictor Variable	Random Forest Importance (Mean Decrease in Gini)	Multiple Regression β (Standardized)	p-value
Assignment submission rate	0.312	0.41	< 0.001
Login frequency	0.287	0.33	< 0.001
Learning resource usage	0.235	0.24	< 0.001
Forum participation	0.166	0.18	0.002

Note. Regression $R^2 = 0.57$, $F(4, 379) = 125.3$, $p < 0.001$. All predictors significantly contributed to the model. Variable importance in Random Forest is normalized to sum to 1.

Assignment submission rate was the single most important predictor in both analyses, accounting for 31% of total importance in the Random Forest model and largest standardized regression coefficient ($\beta = 0.41$). It was found that students who submitted work on-time or early consistently received higher overall grades than those submitting late or failing to submit (and the decline was steep). This result is consistent with previous studies showing that submission timing is a significant predictor of time management and self-regulation (Park & Jo, 2015; You, 2016).

Login frequency (Rank 2: 28.7% in Random Forest) Frequent and regular logins corresponded positively significantly with final grades, supporting the idea that continuous presence in the LMS is a better indicator of engagement and classroom awareness throughout the course (Macfadyen & Dawson, 2010). Interestingly, the standard deviation of the inter-login intervals was a stronger correlate than total login count in and of itself, which is not unlike what Gašević et al. (2016) says consistency is more important than quantity.

Use of learning resources, comprising the time spent on course materials and number of different resources accessed, represented 23.5% in the Random Forest model. Students who interacted more deeply with video lectures, readings and self-assessment quizzes scored better than students that accessed only minimal content. This is in line with the cognitive engagement dimension of self-regulated learning (Kovanović et al., 2015).

Engagement in the forums, while an important predictor of learning outcome scores also had a lower weighting (16.6%) compared to all other variables. While active contributors were more likely to get higher grades, the impact was less pronounced than in the case of other variables. This could indicate that in this specific course, forums were optional and primarily used for clarification rather than as core learning activities, a context-specific factor also identified in previous studies (Wise et al., 2012).

5.3 Visualization of Learning Behavior Patterns

In addition, temporal heatmaps and engagement trajectory plots were used to visualize the learning behavior patterns in conjunction with the quantitative measures. Figure 1 (not depicted in text) shows the weekly student login frequencies split by final performance quartile. Top quartile students (high-achievers) logins remained consistently high from the first week all the way to semester's end, except for a slight dip during midterm periods. In contrast, low-performing students (bottom quartile) logged in erratically, with sporadic outputs early on in the course but higher levels of activity prior to critical deadlines only, followed by dramatic declines post-mid-term. This behavior aligns with the “cramming” pattern linked to low self-regulation (You, 2016).

Time from assignment release to submission was characterized in Kaplan-Meier survival curves for each user type. High-performing students submitted assignments on average 2.3 days in advance of the deadline (SD = 1.1 days), while low-performing students submitted on average 1.2 days after the deadline (SD = 2.4 days). By the third set of assignments (week 5), it was clear there was considerable divergence in submission timing, indicating that early interventions based on previous assignments submission patterns may be piece-worthy.

Forum participation was bimodal: 42% (n=374) of students made no posts in the forums, while 28% (n=245) were considered active forum users (>10 posts). Active participants who posted earlier (around the first two weeks) and maintained their posting over the course received significantly higher grades than those who only posted towards the end of the semester. This positively reinforces the building blocks of pre-critical engagement in online writing due to its ability to support early social integration (Dawson, 2006).

Lastly, scatterplots of total time spent vs the number of resources accessed were used to visualize resource usage patterns. Students who both spent above-average time and accessed a wide variety of resources (i.e., engaged in deep, varied interaction) were consistently members of the high-performing group. Students situated low on the time and low on the variety dimension were over-represented in that low-performing group; see Figures 3(a) and 3(b). For example, a small subset (∼8%) of students who devoted very high time but accessed few resources were able to demonstrate lower performance, which may indicate that inefficient study usage patterns are detectable with combined metrics.

In conclusion, the findings show that predictive models, specifically Random Forest, can accurately predict student academic performance by utilizing accessible LMS behavioral data. The two most influential predictors are the assignment submission rate and login frequency, where temporal visualizations highlight distinct engagement patterns that differentiate succeeding from struggling students. These insights serve as a foundation for the development of early warning systems and targeted intervention regimes to be applied in online learning environments.

6. Discussion

Key words: learning analytics, academic performance prediction, online learning journal. The results show that behavioral signals drawn from Learning Management System (LMS) logs can predict student outcomes with significant accuracy, and that certain engagement behaviors most notably assignment submission patterns and login frequency turn out to be the most robust predictors of academic performance. They also add to the literature on learning analytics and educational data mining while providing practical insights for educators and instructional designers. The paper concludes by comparing the findings to previous research, explaining the results in light of theories of self-regulated learning and discussing implications for teaching practice and online course design.

6.1 Comparison with Previous Research

The prediction accuracy of Random Forest used in this study (87.0%) is not far apart from the findings of previous learning analytics studies. Conijn et al. (2017), studying 17 blended courses through Moodle, noted that Random Forest accuracy ranged between 80% and 89%, as a function of course feature and prediction timing. Similarly, Hellas et al. To put this in perspective, one systematic review of predictive models developed for educational purposes (see Shih et al. The findings of the current study validate these previous results and generalise them to a developing country online education system, a context that has received limited attention in the learning analytics literature (Viberg et al., 2018).

The observation that students' assignment submission rate was the best predictor of performance is consistent with previous literature highlighting the importance of time management and self-regulation in online learning. Park and Jo (2015) found that the timing of students' assignment submissions was one of the most significant predictors of final grades, with students who submitted their assignments early or on time performing best. You (2016) also found that the regularity of assignment submission throughout a course was more predictive than total login count or other indicators of engagement. The current research extends these results by showing that assignment submission patterns are not only associated with outcomes but also exhibited the highest variable importance in machine learning models, indicating an appropriate behavioral trigger for prioritizing inclusion in early warning systems.

The second most important predictor was login frequency, which supports a quite large body of research related to LMS access patterns and academic performance. The extent to which students logged in has been shown to be a significant predictor of student success in terms of course grade in a large undergraduate course (Macfadyen & Dawson, 2010), and was followed by many other studies that confirmed the result (e.g., Morris et al., 2005; Tempelaar et al., 2015). But the current study provides more complexity to this relationship by revealing that regularity of login essentially the degree to which access patterns remained consistent across time was more clearly linked to outcomes than total logins alone. This is in line with Gašević et al. (2016), who contended that aggregate measures of engagement may obscure meaningful temporal dynamics, and that the distribution of engagement over a course matters more than simple totals.

The predictive contribution of learning resource usage (23.5% variable importance) is consistent with previous work that shows deeper interaction with course materials predicts better learning outcomes [13]. Kovanović et al. (2015) highlighted that time-on-task estimates need to be contextualized, where time on meaningful learning tasks is often a better predictor compared to time spent passively consuming content. The finding that both how long and how many kinds of resources accessed contributed to prediction is in line with this more nuanced view, which the present study contributes to. Giannakos et al. (2015) found that video interaction patterns, including completion rates and replay behaviors, predicted performance significantly [14], highlighting the need for studies to include finer granularity of resource interaction going forward.

Forum involvement (while statistically significant) was the least powerful predictor of the four variables evaluated. These findings should be treated with caution, however; previous research has produced consistent but uncertain evidence of whether engagement in discussion forums predicts viewers' ultimate viewing behavior (Katz & Lazarsfeld 1955). Dawson (2006) found that sense of community and persistence were both positively correlated to forum participation in a statistically significant way, though not necessarily with final grades. Wise et al. In 2012 they argued that the amount of contributions to the forums set online communities apart, and therefore by simply counting posts we missed out on understanding what engagement really looks like. This may further explain the relatively weaker predictive power of forums: in our present study, when used at all, forums were mainly as a vehicle for logistical clarifications rather than for ongoing substantive discussion about the course material. This high-level insight also serves as a reminder that course design greatly influences which engagement behaviors may become meaningful predictors of success (Gašević et al., 2016).

6.2 Theoretical Implications: Self-Regulated Learning

These findings add empirical weight behind the application of self-regulated learning theory as an understanding platform for student engagement in online environments. The theory of self-regulated learning suggests that successful learners are those who actively regulate their cognition, motivation, and behavior through forethought, performance control, and self-reflections processes (Zimmerman & Schunk, 2001; Pintrich et al., 2003). The behavioral indicators analyzed in this study (i.e., login frequency, assignment submission patterns, resource usage and forum participation) might be seen as observable manifestations of these underlying self-regulatory processes.

Assignment submission behaviour is indicative of the performance control phase of self-regulation, specifically time management and goal adherence. Students who hand in assignments on time are able to plan their study schedules, keep track of progress against deadlines and modify effort where necessary (Broadbent & Poon, 2015). The strong predictive power of this variable indicates that time management is an important self-regulatory skill in online learning environments, where there is less external structure compared to face-to-face environments. Login frequency (i.e., logged in number of times throughout the course) and regularity capture the forethought and monitoring phases of self-regulation, with students that log in consistently more likely to remember what is due on a given day, decide which things to do when, and stay engaged with their work over time (Winne & Hadwin, 2008). The use of learning resources reflects the cognitive engagement dimension of self-regulation, as students who engage actively and strategically with materials display metacognitive awareness and adaptive learning strategies (Kovanović et al., 2015).

Although this study shows that forum participation has weaker predictive power, this does not undermine the theoretical significance of social engagement to self-regulated learning. It rather indicates that learning activities, by their design, shape which self-regulatory behaviours are most fundamental to learners' success. If a course is designed considering forums as the main part of learning activities, forum participation may become an even stronger predictor (Joksimović et al., 2016). This finding highlights the importance of considering instructional context in interpreting learning analytics results.

6.3 Implications for Teachers

The results of this study provide IMPLICATIONS in practice for teachers, and instructors in online learning environments. First, the high predictive accuracy attained by models using early behavioral indicators suggests that teachers can employ LMS data to identify at-risk students in time for intervention earlier in the semester when interventions are most likely to be accounted. Instead of waiting for midterms to expose academic struggles, educators can keep track of engagement signals (like patterns in assignment submission and log-in frequency) beginning in the first weeks of a course. Research by Arnold and Pistilli (2012) showed that early warning systems based on learning analytics are good at flagging distressed students who might prompt efforts to reach out and help them, so that retention and performance improve as soon as those "at-risk" students can be identified.

Second, the finding that submission patterns of assignments are the best predictor supports understanding student engagement as connected with monitoring submission lateness. Teachers

can access LMS analytics dashboards to observe submission trends across the class and spot students in the batch who are continuously late or missing their assignments. These students can be identified for interventions, such as reminders, check-ins or referrals to academic support services. Crucially, interventions should be presented as supportive rather than punitive to emphasize that help and resources are available in the organization, as opposed to being punished or called out for a deficit (Sclater et al., 2016).

Third, the finding that consistency of login frequency matters more than overall logins suggests teachers should be concerned with engagement patterns rather than raw volume. So a student who logs in 50 times across a semester but only during the week before deadlines is likely to be more at risk than, say, someone who logs in 20 times but consistently throughout the semester. Teachers can investigate temporal visualizations of login activity to pinpoint students whose engagement patterns are erratic so they might reach out and learn about barriers to consistent participation.

Finally, teachers should know the predictive power of engagement indicators might look different in courses that are less prescriptive than their traditional counterparts. In those course types that do not have forums as a feature, or where forums are comparatively peripheral, participation in the forums is not a good predictor of success. By aligning what you monitor with the specific engagement opportunities built into your classes, teachers can have a clearer picture of their students.

6.4 Implications for Online Course Design

The implications of this study also have significant repercussions for the design of online courses. All of this means that the structural features associated with timely submission and success should be carefully considered by those designing new courses hmICON, hn1 and assignment submission patterns might just tell you enough to be able to act. Tips in this area may include the provision of clear and consistent deadlines, breaking larger assignments into smaller, scaffolded tasks, or even automated reminders or scheduling tools which help learners to better manage their time (Broadbent & Poon, 2015). Courses structured around frequent, low-stakes assessments may allow more early detection of struggling students and timely feedback to facilitate self-regulation.

Second, the impact of how often and regularly a student logs in on success suggests that courses should be designed to foster sustained engagement. Instead of expecting students to independently initiate interaction, course designers can design activities that are interspersed throughout the week that require participation (discussions with weekly prompts, knowledge checks with regular feedback loops) and collaborative projects that naturally lead to ongoing communication. Research (You, 2016) suggests that more consistent weekly structured activity in your course leads to more uniform levels of engagement and better outcomes for students.

Finally, the reduced predictive validity found for forum participation in this study implies that course designers need to revisit how discussion forums are used within a learning activity. The when forums are a kind of optional add-on, not something vital to learning, students may not gain enough value from participating. Forums that require substantive contributions, are evaluated for quality and are tightly integrated with course content may help improve engagement as well as the predictive value of data derived from forums (Wise et al., 2012).

Third, consistent with the design of rich and varied resources that allow students to engage hands-on, we found that learning resource usage predicted performance. Multiple formats (e.g., video, text, interactive quizzes), and giving students the chance to track their own progress through materials may support self-regulated learning and offer students opportunities for monitoring their understanding (Kovanović et al., 2015). Moreover, embedding self-assessment tasks that offer immediate feedback enables students to measure their understanding and adapt their study approach accordingly.

6.5 Institutional Considerations

The implications of this study extend not only to individual teachers and course designers, but also to institutional policy and practice. With the increasing maturity of institution-wide learning analytics systems that aggregate data across courses, comprehensive early warning capabilities and coordinated support interventions can be enabled (Sclater et al., 2016). Universities must also invest in the technical infrastructure and human capacity required to implement learning analytics ethically and effectively, including data governance frameworks, faculty training, and support services for students flagged as at risk.'

To summarize, the results from this study corroborate and build on previous work showing significant prediction with LMS behavioral metrics while underscoring assignment submission patterns and login regularity as highly influential predictors. We believe that these findings not only contribute to self-regulated learning theory (SRLT) by providing an empirical basis for understanding student engagement, but they also offer practical insight into how educators, course designers, and institutions might use learning analytics more effectively in ways that support improved student performance in online courses.

7. Implications

These other studies focused on varying aspects and factors contributing to the students engagement have implications for the educational practice, institution policy and wider application of learning analytics in online learning context. This research filled the gap by showing that behavioral indicators captured through LMS can predict with substantial accuracy the academic performance of students, thus providing empirical support to guide data-driven interventions and instructional strategies. The implications rendered in this section are cast in terms of practical implications early warning systems and enhanced pedagogical practices and policy implications implementation of analytical frameworks across institutions.

7.1 Practical Implications

7.1.1 Early Warning Systems for Struggling Students

The most relevant aspect of this study regarding practical implications has to do with the formation and realization of early warning systems using Learning analytics for detecting students at risk of academic failure. The Random Forest model achieved a high predictive accuracy (87.0%), suggesting that machine learning algorithms can be successfully applied to differentiate students who are likely to complete a course from those at risk of dropping out, based on then available data from the first few weeks of the semester. This ability fills a major

gap described by the problem statement, where predictive systems do not exist to identify at-risk students in advance of being able to impact their changes in behavior.

However, early warning systems using learning analytics have some advantages over other techniques of student monitoring. Traditional methods usually depend on assessments like midterm exams, which can happen weeks or even months into a course (Arnold & Pistilli, 2012), and by then students who are having difficulties with the material may have already fallen significantly behind. In contrast, learning analytics allows for real-time monitoring of engagement markers that can foreshadow risk long before formal assessments show problems. As an example, students who do not submit early components of a course/curriculum or have a login frequency below a cut point can be flagged for intervention within the first two to three weeks of even a multi-week course when the ability/contribution to recover is greatest.

The finding that assignment submission patterns are the most predictive of student performance has specific implications for the design of early warning systems. Submission timeliness is a behavioral indicator that is both real-time observable and actionable: when a student misses an assignment deadline or submits late, an early warning system can trigger automated notifications to both the student and the instructor so they can do something about it right away. Park and Jo (2015) found that early intervention with such near-real-time interventions can better improve student outcomes, especially when combined with tailored feedback and support resources.

Nonetheless, any early warning system implementation must account for ethical and logistical challenges. Institutions should develop clear policies on data collection, privacy, and the use of predictive analytics to ensure that students understand how their data is being used and that interventions are supportive rather than punitive (Sclater et al., 2016). Moreover, early-warning systems should work alongside human judgment, not in place of it. The models generated in this study are not designed to supplant educators, but rather to aid them by flagging students who may need enhanced attention; however, decisions about interventions should still rest with instructors and academic advisers that can contextualize analytic results against qualitative observations and student situations.

A second point of consideration relates to the necessity of appropriate response mechanisms when students are flagged as at risk. Supporting services such as academic advising, tutoring, counseling, and accessibility resources must be aware of their newly identified students and should be resourced to assist them (Siemens & Long 2011). An early warning system that identifies risk but does not connect students to support will likely not improve student outcomes, and might even lead to frustration or anxiety among the very students whose well-being an EWS is intended to protect. As a result, early warning system development should come alongside investments in student support infrastructure and training for the individuals responding to alert signals.

7.1.2 Improved Online Teaching Strategies

The results of this research will also guide the development of enhanced online teaching practices by clarifying what engagement behaviours have the highest impact on student success. The finding that assignment submission rate was the strongest predictor indicates that instructors should focus on encouraging assignment completion and time management strategies. Such support may take the form of breaking larger assignments down into smaller, scaffolded tasks

with interim deadlines, providing clear rubrics and expectations for performance, or using automated reminders to keep students on track (Broadbent & Poon, 2015). Instructors can also use LMS analytics to track submission patterns across the class and flag students who may be struggling with time management before they fall too far behind.

The result that login regularity (ie, the consistency of access patterns) is more important than total login count can inform how instructors think about and encourage engagement. Instead of simply motivating students to log in multiple times per week, faculty should emphasize the need for regular, consistent engagement across the entire semester. By incorporating weekly activities, consistent feedback loop and predictable rhythms of interaction into course designs (You, 2016), they also might help to encourage more steady engagement patterns. Since instructors serve as the main authority figure in their course a setting, they can model desired behaviors by showing up consistently to their course, responding with timely feedback on student questions, and discussing why attendance is important throughout the term.

Where the usage of learning resources are a crucial aspect for prediction performance, instructors should consider creating classes with rich and diversified learning materials where students can actively interact while using the course. Offering various formats for content delivery such as video lectures, interactive modules, readings and self-assessment quizzes can meet diverse learning preferences and enhance engagement (Kovanović et al., 2015). Educators can also use analytics to determine how students engage with resources and identify materials that may be particularly difficult or engaging, leading to iterative improvement of courses.

This study highlights the importance of intentional design for social learning activities, as evidenced by the weaker predictive power of forum participation than in other studies. In cases when forums are optional or marginal, students do not receive enough benefit from being involved. To use discussion forums as impactful learning events, instructors can design effective forums that require meaningful contributions to the thread, assess forum participation and set expectations regarding user experience in terms of quality and volume (Wise et al., 2012). Instructors can also model how to participate effectively in forums and give feedback promptly, so that students will become involved more continuously.

Lastly, the background theory underlying this study traces its roots from self-regulated learning theory which implies that it is possible for instructors to promote students together with effective performance through explicit instruction in self-regulation strategies. This could involve instruction on goal setting, time management, metacognitive monitoring and help-seeking behaviors (Zimmerman 2002). The risk of students being unable to succeed in online learning environments may be mitigated by embedding prompts for self-regulation into course activities, e.g. tasks that involve pre-assignment planning exercises or reflect on questions after an activity is completed above and beyond the academic skills required.

7.2 Policy Implications

7.2.1 Institutional Adoption of Learning Analytics

These findings contribute to our understanding of the factors affecting adoption and implementation of learning analytics systems, with important implications for institutional policy. With the growth of digital educational infrastructure, particularly for teaching and learning, institutions will be challenged to extract actionable insights from the data being

generated as a fundamental step in realizing online education's potential (Siemens & Long, 2011). The predictive accuracy shown by LMS behavioral data models presents a strong rationale for institutional efforts to move toward systematic approaches for learning analytics that go beyond isolated pilot projects and do embrace institution-wide frameworks.

Broad based institutional adoption of learning analytics requires a robust data governance framework regarding ownership, privacy, security and ethical use of the data. Students should be made aware of the data that is being collected about them, how it is being used and with whom (Sclater et al., 2016). Ensuring Predictive Models Support Students, Rather Than Penalty ThemInstitutions need to articulate clear protocols for obtaining informed consent, anonymizing data for research purposes, and allowing predictive models tp serve as support instead of penalization mechanisms to the students. This also needs to take into account issues of algorithmic fairness and bias, with the goal being that predictive models do not disproportionately flag students from particular demographic groups or propagate existing inequities (Ferguson, 2012)

A second policy consideration is the relationship of learning analytics to current institutional systems and processes. Effective early warning systems will only work for the institutions they serve when there is coordination between academic departments, student support services, information technology and institutional research units. Institutions need to create concrete pathways for how alerts generated by analytics will be acted upon- this is where we can ensure that students of concern are helped in a timely and effective fashion (Arnold & Pistilli, 2012) FAQs: This might involve rethinking traditional roles and responsibilities, creating new channels for communication, and investing in staff training to build capacity for data-informed practice.

Institutional policy should turn to the sustainability of learning analytics initiatives. Ongoing investment in technical infrastructure, data management and analytical expertise is required to develop and maintain predictive models. This means that institutions should think about how they can distribute resources across the entire learning analytics lifecycle from data collection and model building, to implementation, evaluation and iteration (Lang et al., 2017). This can take the form of building specialised learning analytics teams, working with third-party vendors or developing academic expertise in data analytics through training and development initiatives.

Finally, organizations need institutional policy to support research and evaluation efforts that help advance the evidence base for learning analytics. Although this study shows that behavioral indicators may have predictive power in a particular course context, additional research is needed to study whether findings generalize across courses, disciplines, and institutional contexts. Institutions must develop processes to allow for ethical access to learning analytics data, encourage faculty engagement with learning analytics research, and disseminate research findings in order to contribute meaningfully to the wider educational community (Viberg et al., 2018).

7.3 Summary

This study has implications for various levels of education practice and policy. Ultimately, these findings lend support to developing ears to the ground through early warning systems for identifying at-risk students in advance of meaningful intervention and more strategic teaching practices focused on meeting students where they are in terms of assignment submission,

engagement with coursework, and improving course design. The results indicate at the policy-level provide a case for institutions adopting learning analytics frameworks, and be mindful of data governance, integration with student support services, sustainability and research capacity. When taken together, these implications suggest a future where learning analytics are systematically activated to help students succeed in online distributed learning.

8. Limitations

Although this study provides important insights into the use of learning analytics for predicting student performance in online learning contexts, it is also necessary to recognise several limitations that limit the generalisability and scope of these findings. Understanding these limitations is important for appropriately interpreting the results and for guiding future research efforts that entail bridging the key gaps identified. The limitations outlined in this section relate primarily to the data source and the variables we considered, as well as additional considerations relating to sample characteristics and methodological limitations.

8.1 Data Limited to One Institution

One of the major limitations of this study is that it has only collected data from one institution, namely organisation in developing country context students at one university. Though this focus provided detailed analysis and control over institutional variables, it severely limits the generalizability of what we found across other types of educational contexts. Research into learning analytics has shown that predictive models designed in one context cannot be used directly in another because if the differences, such as student populations, institutional policies, courses designs or technology infrastructures (Gašević et al., 2016). What drives success is not a fixed formula, but rather bases itself on institutional differences and dynamics and even the metrics for success with academic performance being deep down some times based in dealing by proxy as comparative.

The single-institution design limitation is especially pronounced given that most of the learning analytics research to date has been conducted in developed countries, specifically North America, Western Europe and Australia (Viberg et al., 2018). Although the current paper offers badly needed evidence from a developing country context, findings from one institution cannot be expected to capture the diversity of online learning environments in multiple regions, cultural and educational systems. Technological infrastructure, the reliability of internet access and levels of digital literacy often differ across contexts (Kennedy et al. 2013; Hurd 2004) and are likely to affect rates of engagement as well as the extent to which behavioral indicators can be used for predicting purposes.

In addition, this study concentrated on one course at the institution; although this permitted control of lower-level variables such as instructional design and assessment design, it limits generalizability to other disciplines, levels of courses and types of instruction. A predictive model that you have built for a certain course may not generalize to another course with a different type of content, teaching style, or student population (Conijn et al., 2017). The particular course that was examined may have had specific features (e.g., discussions forums are optional, the frequency of assessments, etc.) that influenced which behaviors turned out to be their strong predictors.

Including a sample of n = 384 that is sufficient for the analysis procedures from this study also pertains as a limitation. Larger, more diverse samples would provide more statistical power and enable the use of more complex modeling approaches while providing greater confidence in findings as stable and generalizable. The sample was drawn from a single-semester cohort and may not reflect potential between-academic-term or between-cohorts differences. Robustness of the predictive models would also require replication across multiple semesters and courses.

8.2 Limited Variables

The second important limitation relates to the variables considered in the analysis. The study features four behavioral indicators extracted from LMS logs: login rate, assignment submission rate, forum contribute and learning resource usage. Though these variables constitute widely studied predictors of academic performance used in learning analytics (Macfadyen & Dawson, 2010; Park & Jo, 2015), they only account for a fraction of the drivers of students' academic outcomes. These models may be less satisfactory due to the omission of variables that could provide better explanatory power, thus potentially leading to omitted variable bias.

Importantly, the analysis did not control for demographic factors such as age, gender, precollegiate academic performance, socioeconomic status or educational background. These factors have been consistently shown by research to relate to differences in student outcomes and can interact with engagement behaviours in complex ways (Tempelaar et al., 2015). The study also cannot tell whether the predictive power of behavioral indicators is independent of demographic characteristics, or if certain demographic groups are systematically flagged by the models. This concern highlights a number of significant questions related to algorithmic concepts of fair outcomes and the role predictive models can play in entrenching or exacerbating inequities already present in educational systems (Ferguson, 2012).

The study also omitted measures of student motivation, self-efficacy, and other psychological constructs that the authors argue are theoretically relevant to self-regulated learning. Although the behavioral markers considered here are viewed as observable manifestations of self-regulation, the lack of direct assessments of motivation or metacognition prevents confirmation of the theoretical framework behind it (Zimmerman, 2002). Overall, future research might utilize surveys in conjunction with behavior-oriented data to provide a rich view of students that matters for their success.

The study also missed collecting data on the quality of engagement, only concentrating on quantitative measures. For instance, forum participation was recorded in terms of number of posts but not the depth or quality of contributions. Similar to learning resources usage was quantified by time and number of resources accessed but not by nature of interaction with those resources (e.g., replay patterns, notes, annotation). Research by Wise et al. (2012) and Giannakos et al. (2015) that qualitative dimensions of engagement are often more predictive than naive quantitative measures and so the current models may be underestimating the predictive utility of engagement data.

The study also did not aggregate institutional data, including academic advising records, library usage, or participation in support services. These variables may also help to provide context for understanding student success, and as such, they are potential moderators of the engagement-

outcomes relationship. However, such data are lacking which limits the ability to make operational recommendations for institutional mechanisms of support.

8.3 Additional Methodological Limitations

In addition to limitations related to data source and variables, various methodological limitations deserve acknowledgment. The study used a retrospective design and analyzed data from an already completed course to make predictive models. Although this perspective is typical of learning analytics literature, it misses real-time intervention and truly dynamic characteristics. Experimental designs that evaluate predictive models in real time with actual interventions are necessary to confirm the causal efficacy of learning analytics upon student performance (Sclater et al., 2016).

The binary classification of academic performance (pass/fail) is also a limitation, as it diminishes the richness of continuous-grade data and may hide variations in results among students passing the course. Going forward, more nuanced relationships of interest could be captured by regression-based approaches forthcoming between engagement and performance outcomes such as continuous target behaviors like social interactions directly observed or hitting predetermined benchmarks.

First, the study did not consider changes in uptake of engagement behaviors that may have occurred due to the ongoing COVID-19 pandemic; the data were collected during a period characterized by students facing disruptions to their learning environments. It is not clear how much these findings reflect normal online learning circumstances rather than pandemic-induced ones, and replication under more stable conditions would be useful.

8.4 Final Statement

In conclusion, this study has limitations related to a single institutional and course dataset and few behavioral variables. Such limitations affect the generalizability of findings and suggest a need for future studies that encompass different institutional contexts, broader ranges of variables, and prospective research designs. Recognising these limitations does not subtract from the study's contributions but rather places its findings in context and highlights fruitful avenues for future research.

9. Future Research Directions

Though this study has validated the usefulness of learning analytics for predicting student performance in online spaces, limitations and findings lead to suggestions for future research directions. We will articulate how future research in learning analytics should focus on addressing the challenges of single-institution studies, broadening the spectrum of variables under investigation and investigating the interplay between predictive analytics and adaptive learning technologies. The subsequent sections present critical lines for future research that can complement the groundwork set forth in this study.

9.1 Cross-Institution Datasets

A key avenue for future work is collecting and analyzing cross-institution datasets that will allow researchers to explore the diversity of online learning contexts across institutional types,

geographic regions, and student populations. As with most of the existing learning analytics literature, the current study was constrained to one institution, and therefore raises questions regarding whether or not findings can be generalized to other educational context (Viberg et al., 2018). Future work should be controlled by the establishment of large multi-institutional datasets that allow us to investigate whether an individual falls into damaging third or not, on a larger scale context and whether all behavioral markers have unit predictive power in any context (global predictiveness) or they may vary between persons.

The creation of cross-institution datasets would enable several important lines of inquiry. First, researchers might examine how well predictive models developed using data from one institution are transferred to different institutions. There are important practical implications to this question because institutions exploring the adoption of learning analytics often do not have historical data available at an institutional level to build models (Gašević et al., 2016). If models show generalizability across institutions, it may obviate the need for duplicative work within insititutions and facilitate adoption of available models from other sites. Relatedly, if models are heavily context-specific this will reflect the necessity for each institution to build its own predictive models, or more complex domain adaptation procedures.

Second, cross-institution datasets would allow for investigation of how institutional characteristics including size, selectivity, public vs. private status and resource availability moderate the association between engagement behaviors and academic outcomes. These analyses could help determine if certain engagement indicators are more or less predictive in various institutional contexts and help target interventions accordingly. Also, research between institutions could explore if predictive models demonstrate differential performance according to student demographic groups addressing pressing questions of algorithmic fairness (Ferguson, 2012).

Third, comparative studies across countries and cultural contexts would answer the research gap originally identified in this paper (that learning analytics is under-studied in developing countries relative to traditional parts of the world). Subsequent investigations should proactively integrate institutions from different geographic contexts, such as those in Africa, Asia, Latin America and the Middle East to ascertain how cultural context; educational traditions; and technological underpinnings influence engagement trends and predictive value of behavioral indicators (Tlili et al., 2020). Such studies would help broaden access to learning analytics globally and assist designing interventions that are culturally responsive.

Cross-institution datasets are high-demand products that would require a lot of coordination and collaboration between institutions as well as significant attention to data governance, privacy and ethical considerations. Building these artifact accounts would require increased collaboration between researchers, including initiatives like the Society for Learning Analytics Research (SoLAR) (Lang et al., 2017), and shared data repositories and data standards. Funding agencies and educational organizations can play a catalytic role by providing support for the infrastructure necessary to conduct learning analytics research at multiple institutions.

9.2 AI-Based Adaptive Learning Systems

A second key direction for future work involves the amalgamation of predictive analytics with AI-based adaptive learning systems capable of delivering personalized, real-time interventions

based on patterns of student engagement. This study shows that predictive models can identify at-risk students, but it does not speak to how such predictions might inform practice. Future work should investigate the design and assessment of adaptive learning systems that are informed by predictive analytics and use students' engagement and performance data to make real-time adjustments to their learning experience (Siemens & Baker, 2012).

AI-based adaptive learning systems are a departure from traditional modes of instruction, which can be termed as a ‘one-size-fits-all’ approach on towards personalized learning experiences that respond to individual student needs. These systems can include various types of artificial intelligence, including machine learning for student modeling, natural language processing for analyzing students' answers, and recommender systems to recommend learning resources and activities (Koedinger et al., 2015). In conjunction with predictive analytics, these systems may then automatically detect learners with behaviors similar to others at risk and recommend specific support such as personalized feedback, recommended study strategies, guided learning paths or linkages to supportive resources.

There are a number of specific research questions that deserve exploration in this area. First, which types of AI-driven interventions are most effective for at-risk students? Research could compare alternative intervention strategies such as personalized messages, dynamic ordering of content-based engagements, or recommendations for peer study groups to identify the approaches resulting in maximally improved engagement and outcomes. Second, how do adaptive systems balance automation and human oversight? Although automated interventions can help deliver timely support at scale, there are likely important roles for instructors and advisors in acting on analytics alerts. Further studies should also examine hybrid models that integrate AI-facilitated assistance and human discernment, together with relationship-centered interventions (for example Arnold & Pistilli 2012).

Third, which quality dimensions of engagement cannot be captured by simple behavioral metrics, and how can adaptive learning systems leverage such qualities? As discussed in our study’s limitations, quantitative measures like frequency of log ins or number of posts do not always reflect the depth or quality of student engagement. Future studies should consider adopting natural language processing to analyze forum contributions, machine learning techniques for classifying resource interaction practices and computer vision modeling approaches for video viewing behavior (Giannakos et al., 2015). Incorporating these richer engagement measures could provide predictive power that models with simplistic jois-only covariates miss, and encourage more nuanced interaction interventions.

4 Long-term follow-up studies are required to keep tabs on lasting influences of AI-based adaptive learning systems on student achievements. The majority of studies to date have looked at outcomes after just one semester, leaving the question still unanswered of whether such systems produce enduring gains in student learning, persistence and the development of self-regulated learning skills. Future research should follow individual students across multiple courses and academic terms to better understand the cumulative effects of adaptive interventions, and to investigate whether improvements transfer to different learning contexts.

9.3 Additional Research Directions

In addition to cross-institution datasets and AI-based adaptive systems, there are different research directions worth considering. First, future studies should include more variables than those used in the current study. This includes demographic variables to analyze algorithmic fairness, psychological measures such as motivation and self-efficacy to fully test self-regulated learning theory, and qualitative engagement metrics to capture the depth of student interaction (Wise et al., 2012). Mixed-methods research that strategically combines quantitative analytics with qualitative insights derived from student interviews or instructor observations could deepen our understanding of the mechanisms by which engagement is linked to outcomes.

Second, future studies should use prospective, longitudinal designs that test predictive models in real time with actual interventions. Retrospective analyses are instrumental for developing models, however the true test of learning analytics effectiveness depends on whether a predictive model can be applied to improving student outcomes through timely and focused interventions (Sclater et al., 2016). Causal evidence of effectiveness will require randomized controlled trials and quasi-experimental designs comparing outcomes for students receiving analytics-based interventions with appropriate control groups.

Third, future research needs to be more systematic in exploring the ethical dimensions of learning analytics implementation. As predictive models grow more sophisticated and are adopted by institutions in their decision-making, questions of data privacy, algorithmic transparency, student agency and potential bias have become increasingly urgent. Work that engages with these ethical issues and develops frameworks for responsible learning analytics practice is critical to ensuring the field develops in ways which benefit students and sustain trust (Ferguson, 2012).

In conclusion, it must be handled that the future of learning analytics research involves moving from single-institution studies to cross-institutional collaboration, merging predictive analytics with AI-based adaptive learning systems, and tackling the moral and methodological issues which come as the field develops. If pursued, these directions can push forward theory as well as practice around the application of analytics to foster student engagement and ultimately success in online learning environments.

10. Conclusion

The intrinsic potential of Learning Management Systems in a learning environment offers unique opportunities for demonstrating students' engagement characteristics, as it rapidly expands. The work presented in this study has shown that self-regulated learning-based learning analytics can be effectively leveraged to predict academic performance of students with behavioral indicators readily extractable from the LMS log data. The present study similarly advances the use of predictive models in learning analytics by developing and validating models which exploit login frequency, assignment submission, forum activity and learning resource behavior as predictors for improved student outcomes.

Learning analytics is extremely important in today's world. With the growth of digital learning infrastructure in educational institutions comes an increasing volume of data produced from these systems, making it imperative to extract actionable insights as part of realizing the full potential that online education has to offer (Siemens & Long, 2011). Learning analytics describes

a maturation of the educational technology practice, from deploying digital (learning) tools to generate and collect data that reflects how learners engage with them, coupled to a use of these data for understanding how engagements relate to learning outcomes (Gašević et al., 2015). Scientific accomplishment The discipline fills a major void in online schooling the technology to deliver early alerts that can identify at-risk students before academic failure is inevitable. However, traditional methods for monitoring student progress which involve leaning on periodic assessments and instructor observations are insufficient due to decreased visibility into students' learning processes in online contexts that can lead intervention efforts to be implemented only when it is too late (Macfadyen & Dawson, 2010).

This study identified key predictors of student performance which provide empirical guidance for educators and instructional designers directing their efforts in this space. The assignment submission rate was the highest predictive indicator of academic success, pinpointing that time management and adherence to goals are critical aspects in self-regulated learning (Zimmerman, 2002). Students who submitted an assignment on time or early consistently outperformed those who submitted late or after the deadline, hinting that submission pattern tracking should lie at the heart of early warning systems.[1] Frequency of login was second on importance, with frequency of access rated as more predictive than simple counts of logins. This means sustained, consistent engagement across a course matters more than these bursts of activity. For example, how students use the learning resources was a considerable contributor to the prediction accuracy, which supports the interpretation of active and strategic interaction with course materials as reflecting cognitive and metacognitive aspects of self-regulated learning (Kovanović et al., 2015). Of the predictor variables, forum participation was by far the weakest statistically significant predictor, pointing to the course design itself as a key factor in what engagement behaviors serve as indicators of success.

Learning analytics can offer many desired outcomes in online education. On a practical level, the high predictive accuracy achieved on this study lays the foundation for generating warning systems that can help identify at-risk students in time to provide adequate intervention. These systems might flag students who display concerning engagement patterns missing assignments, infrequent logins, limited access of resources as reasons to solicit timely outreach by instructors, advisors or support services (Arnold & Pistilli, 2012). Early intervention is crucial in online learning situations, especially when students can simply drift away unnoticed until they are immeasurably behind. The finding that assignment submission patterns are the strongest predictor suggests that monitoring submission timeliness from the first weeks of a course can give actionable intelligence for intervention.

Outside of early warning systems, learning analytics provides insights for improved online pedagogies and course designs. Analytics can also be used to understand which engagement behaviors are most predictive of success in their courses, allowing Instructors to adjust the way they teach. Applying strategies that promote login on a regular basis and turning assignments in on time, or how activities may be structured to encourage students to work more regimentedly in their study can foster learning skills (Broadbent & Poon, 2015). Designers of courses can use analytics to assess the effectiveness of diverse pedagogical strategies and utilize evidence from data for decisions regarding course structure, assessment design, and activity design as well as pacing.

On an institutional scene, frameworks of learning analytics encourage data-informed decision-making and strategic enhancements in student outcomes. By creating a multidimensional early warning system, institutions that leverage learning analytics infrastructure will be able to coordinate diverse support services across campus and assess the effectiveness of interventions at scale (Sclater et al., 2016). Building of such systems follow data governance, privacy and ethics around theuse of these analytics leveraging insights into existing institutional processes. Appropriately applied, learning analytics can improve institutional capacity to support student success while preserving trust and transparency.

This study also emphasizes the need for context in learning analytics research. The fact that forum participation was a less significant predictor than the other variables seems to be partly due to course-specific factors, such as the optional nature of its discussion forums. This reflection reinforces the importance of institutions developing risk models and prevention strategies based upon their specific population. The limitations of this study particularly the singleinstitution focus and narrow variable specification point to important directions for future research, including cross-institution datasets, integration of AI-based adaptive learning systems with these predictor algorithms, as well as evaluation of algorithm fairness and ethical considerations.

Ultimately, learning analytics is a fundamental way for you to understand and help figure out student success in your online learning experience. Transcending the entire learning management system environment, utilizing behavioral data produced in routine LMS interactions to facilitate predictive models that can spot and identify at-risk students, tailor teaching strategies for improvement, and promote evidence based decision making in general, is one way they are contributing to the learning of these best practices. These findings show that the strongest behavioral predictors for predicting academic performance were assignment submission patterns and frequency of logins on a regular basis, suggesting practical implications for educators, instructional designers. The need for proper use of educational analytics will only grow as e-learning spreads across the globe, to help every student reach their potential. The potential benefits early intervention for struggling students, improved instructional quality and capacity on the part of institutions to support student success position learning analytics as a prominent research and practice priority across digital education.

11. References

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270. https://doi.org/10.1145/2330601.2330666

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, 50(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61–75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, 27, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Coates, H., James, R., & Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary Education and Management*, 11(1), 19–36. https://doi.org/10.1080/13583883.2005.9967137

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17–29. https://doi.org/10.1109/TLT.2016.2616312

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17–29. https://doi.org/10.1109/TLT.2016.2616312

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Dawson, S. (2006). A study of the relationship between student communication interaction and sense of community. *The Internet and Higher Education*, 9(3), 153–162. https://doi.org/10.1016/j.iheduc.2006.06.007

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. https://doi.org/10.1177/0047239520934018

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5–6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Giannakos, M. N., Chorianopoulos, K., & Chrisochoides, N. (2015). Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. *International Review of Research in Open and Distributed Learning*, 16(1), 260–283. https://doi.org/10.19173/irrodl.v16i1.1976

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Hart, C. (2012). Factors associated with student persistence in an online program of study: A review of the literature. *Journal of Interactive Online Learning*, 11(1), 19–42.

Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., & Liao, S. N. (2018). Predicting academic performance: A systematic literature review. *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, 175–199. https://doi.org/10.1145/3293881.3295783

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, 90, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*, 27, 1–12.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781118548387

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Januszewski, A., & Molenda, M. (Eds.). (2008). *Educational technology: A definition with commentary*. Routledge.

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2016). Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education*, 102, 154–167. https://doi.org/10.1016/j.compedu.2016.09.002

Kim, J., Li, S., & Bonk, C. J. (2014). Factors affecting online learner success: A cross-institutional study. *Journal of Interactive Learning Research*, 25(4), 549–573. https://www.learntechlib.org/primary/p/147280/

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333–353. https://doi.org/10.1002/wcs.1350

Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., & Baker, R. S. (2015). Does time-on-task estimation matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics*, 2(3), 81–110. https://doi.org/10.18608/jla.2015.23.6

Kuh, G. D. (2009). The national survey of student engagement: Conceptual and empirical foundations. *New Directions for Institutional Research*, 2009(141), 5–20. https://doi.org/10.1002/ir.283

Lang, C., Siemens, G., Wise, A., & Gašević, D. (Eds.). (2017). *Handbook of learning analytics*. Society for Learning Analytics Research. https://doi.org/10.18608/hla17

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, 59(5), 593–618. https://doi.org/10.1007/s11423-010-9177-y

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). SAGE Publications.

Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, 8(3), 221–231. https://doi.org/10.1016/j.iheduc.2005.06.009

Murphy, E., & Rodríguez-Manzanares, M. A. (2009). Teachers' perspectives on motivation in high-school distance education. *Journal of Distance Education*, 23(3), 1–24. https://www.ijede.ca/index.php/jde/article/view/543

Osmanbegović, E., & Suljić, M. (2012). Data mining approach for predicting student performance. *Economic Review*, 10(1), 3–12. https://www.ef.untz.ba/wp-content/uploads/2017/06/ER_2012_10_1_osmanbegovic.pdf

Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science*, 21(1), 110–133. https://doi.org/10.3217/jucs-021-01-0110

Patterson, B., & McFadden, C. (2009). Attrition in online and campus degree programs. *Online Journal of Distance Learning Administration*, 12(2), 1–8.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Powers, D. M. W. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63. https://arxiv.org/abs/2010.16061

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. https://doi.org/10.1007/BF00116251

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355. https://doi.org/10.1002/widm.1355

Sclater, N., Peasgood, A., & Mullan, J. (2016). *Learning analytics in higher education: A review of UK and international practice*. Jisc. https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education

Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254. https://doi.org/10.1145/2330601.2330661

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, 46(5), 30–32. https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson Education.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47, 157–167. https://doi.org/10.1016/j.chb.2014.05.038

Tlili, A., Zhang, J., Papamitsiou, Z., Manske, S., Huang, R., Kinshuk, & Hoppe, H. U. (2020). Towards utilising emerging technologies to address the challenges of using Open Educational Resources: A vision of the future. *Educational Technology Research and Development*, 68(2), 789–809. https://doi.org/10.1007/s11423-019-09732-4

Turnbull, D., Chugh, R., & Luck, J. (2021). Learning management systems: A review of the research literature. *Journal of Information Technology Education: Research*, 20, 99–121. https://doi.org/10.28945/4689

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Lawrence Erlbaum Associates.

Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y. T. (2012). Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. *Instructional Science*, 41(2), 323–343. https://doi.org/10.1007/s11251-012-9230-9

World Economic Forum. (2021). *The future of online learning: How education technology is reshaping global education*. World Economic Forum.

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, 29, 23–30. https://doi.org/10.1016/j.iheduc.2015.11.003

Yu, T., & Jo, I. H. (2014). Educational technology approach toward learning analytics: Relationship between student online learning behavior and academic performance. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge*, 269–270. https://doi.org/10.1145/2567574.2567594

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2