

Article

Not peer-reviewed version

Learning Analytics for Predicting Student Performance in Online Learning Environments

[Sayed Mahbub Hasan Amiri](#)*

Posted Date: 17 March 2026

doi: 10.20944/preprints202603.1312.v1

Keywords: learning analytics; machine learning; online learning environments; student engagement; student performance prediction



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

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 growth of online education has led to an increasing amount of data in Learning Management Systems, but many organizations have yet to find ways to effectively use this data to assist students. In this work, we investigate the possibility of using learning analytics for predicting students' academic performance and drop-out in digital learning environments. Utilizing interaction data from university students on a platform, e.g. Moodle or Canvas, during a semester, we build and compare several predictive models, including logistic regression, decision trees, and random forests. We show that certain specific, quantifiable activities (such as login frequency, submitting assignments on time, participation in discussion forums) are the best predictors of final grade, or at least final grade after having dropped out. The random forest method was the best performing one, indicating that it is a robust method to be used for early detection of students who may require additional help. These results show the possibility of learning analytics to go beyond rudimentary descriptive statistics and towards actionable insights. The implications are practical and powerful: institutions may deploy early warning and timely intervention services and educators may understand more clearly which engagement patterns promote success and are more 'responsive' to course design. In the end, this study highlights the importance of moving from intuition-driven to data-driven teaching approaches to not only make online learning more accessible but also more equitable and effective. By targeting on behavioral data that is currently collected, we may as well lay down the first building block for more individualized, supportive digital learning experiences.

Keywords: learning analytics; machine learning; online learning environments; student engagement; student performance prediction

1. Introduction

1.1. Background of the Study

The landscape of education has undergone one of the most profound transformations in its history. Over the past decade, and particularly accelerated by the global pandemic, online learning and digital education have experienced unprecedented growth, fundamentally reshaping how students learn, how educators teach, and how institutions conceptualize the delivery of education (Liu & Yu, 2023). The global e-learning market, valued at approximately USD 197.2 billion in 2020, is projected to reach nearly USD 494.8 billion by 2026, reflecting a compound annual growth rate of over 16.7% (Stratview Research, 2026). This remarkable expansion signals not merely a temporary shift but a permanent evolution in the educational paradigm, one where digital platform have become central to the teaching and learning experience.

At the heart of this digital transformation lie Learning Management Systems (LMS), which serve as the technological backbone for online and blended learning environments. Platforms such as Moodle, Canvas, Blackboard, and Google Classroom have become ubiquitous across educational institutions worldwide, facilitating course delivery, content distribution, assessment administration, and perhaps most importantly, capturing vast quantities of data about how students engage with their learning (Alhothali et al., 2022). These systems generate detailed digital footprints of student

activity every login, every resource viewed, every discussion post submitted, and every assignment uploaded creates a rich tapestry of behavioral data that was simply unavailable in traditional classroom settings (Qiu et al., 2022). As Hubbard and Amponsah (2025) note, the prevalence, presentation, and near real-time nature of the data available through digital learning platforms enable unprecedented insights into the learning process.

This explosion of educational data has given rise to the field of Learning Analytics, defined as the measurement, collection, analysis, and reporting of data about learners and their contexts for the purposes of understanding and optimizing learning and the environments in which it occurs (Alalawi et al., 2024). Learning Analytics sits at the intersection of educational technology, data science, and pedagogical research, offering the potential to move beyond intuition-based teaching toward data-informed strategies that can genuinely enhance student outcomes (Wu, 2026). By systematically analyzing the behavioral traces students leave behind as they navigate online learning environments, researchers and educators can begin to identify patterns, understand engagement dynamics, and ultimately predict which students are on trajectories toward success and which may be heading for academic difficulty.

The importance of this endeavor cannot be overstated. Contemporary higher education faces persistent challenges around student retention, completion rates, and equitable outcomes. Traditional approaches to identifying struggling students often rely on reactive measures poor performance on midterm examinations or failing grades on major assignments by which point intervention may come too late (Moreno-Marcos et al., 2025). The digital nature of modern learning environments, however, offers the possibility of early, proactive identification based on fine-grained behavioral indicators that precede academic decline. As Alalawi and colleagues (2024) argue, the integration of machine learning and pedagogical approaches for student performance prediction and intervention represents one of the most promising frontiers in educational technology research.

1.2. Problem Statement

Despite the transformative potential of online learning, a persistent and troubling reality remains: many online learners struggle academically. The flexibility and accessibility that make digital education attractive can also become sources of difficulty, as students must navigate learning experiences without the structural support, immediate feedback, and social presence of traditional classrooms (Clarín & Baluyos, 2022). Research has documented elevated rates of attrition in online courses compared to face-to-face equivalents, with students often disengaging gradually and imperceptibly until they reach a point of no return academically (Bashiru & Malgwi, 2026). The challenges are particularly acute in contexts where students face additional barriers limited digital literacy, unreliable internet connectivity, competing work and family responsibilities, or insufficient self-regulation skills (Clarín & Baluyos, 2022).

Compounding these individual-level challenges is a systemic one: most educational institutions lack robust predictive systems to identify at-risk students early enough to mount effective interventions. While Learning Management Systems diligently collect vast quantities of data, this information often remains underutilized, sitting in databases as dormant records rather than being transformed into actionable intelligence (Hubbard & Amponsah, 2025). Faculty members, even those deeply committed to student success, typically lack the tools to monitor hundreds of students' engagement patterns continuously and to detect subtle signs of disengagement before they manifest as course failure. As Hubbard and Amponsah (2025) observe, institutions produce terabytes of data generated by thousands of people, yet the gap between data collection and meaningful intervention remains stubbornly wide.

This problem is particularly acute in developing country contexts, where resource constraints, large class sizes, and limited academic support systems compound the challenges of online education (Bashiru & Malgwi, 2026). Research from institutions in Nigeria, for example, has documented that delayed interventions often contribute to poor academic outcomes and increased dropout rates, with early warning systems remaining the exception rather than the norm (Bashiru & Malgwi, 2026). Even

in well-resourced educational systems, the translation of learning analytics research into practical, scalable early warning systems has proceeded slowly, leaving many students to navigate online learning environments without the safety net that predictive analytics could provide.

1.3. Research Objectives

In response to these challenges, this study pursues three interconnected objectives. First, it aims to systematically analyze student interaction data generated within online learning platforms, examining the digital traces that learners leave as they engage with course materials, assessments, and peers. By collecting and preprocessing log data from Learning Management Systems such as Moodle or Canvas, this research seeks to transform raw activity records into meaningful behavioral indicators that can shed light on the learning process (Qiu et al., 2022).

Second, this study endeavors to develop and validate predictive models capable of forecasting student academic performance based on their online learning behaviors. Drawing on established machine learning techniques including logistic regression, decision trees, random forests, and support vector machines the research will compare the accuracy, precision, recall, and F1-scores of alternative modeling approaches to identify those best suited to the educational context (Alhothali et al., 2022). The goal is not merely to achieve statistical prediction but to generate models that are interpretable, actionable, and transferable across different course contexts.

Third, this research seeks to identify the specific learning behaviors and engagement patterns that most strongly influence academic performance in online environments. While predictive accuracy is valuable, understanding which behaviors matter most and how they matter is essential for designing effective interventions and informing pedagogical practice (Alalawi et al., 2024). By examining variables such as login frequency, assignment submission timing, discussion forum participation, video lecture viewing, and time spent on learning resources, this study aims to move beyond black-box prediction toward genuine insight into the dynamics of online learning success.

1.4. Research Questions

To guide this investigation, three primary research questions have been formulated:

RQ1: Which learning behaviors predict student success in online learning environments? This question seeks to identify the specific, measurable actions and engagement patterns that distinguish successful students from those who struggle. Drawing on theoretical frameworks such as self-regulated learning theory, it examines how behaviors like regular platform access, timely submission of assignments, active participation in discussion forums, and consistent engagement with learning resources contribute to academic achievement (Qiu et al., 2022; Wu, 2026).

RQ2: How accurately can learning analytics models predict academic performance in online courses? This question addresses the technical dimension of the study, evaluating the predictive power of alternative machine learning algorithms and feature sets. It considers not only overall accuracy but also the trade-offs between different performance metrics and the practical utility of predictions for early intervention (Alhothali et al., 2022; Moreno-Marcos et al., 2025).

RQ3: What factors influence student engagement in online learning environments? Recognizing that engagement itself is a complex, multi-faceted construct, this question explores the individual, contextual, and instructional factors that shape how and why students engage (or fail to engage) with online learning. It considers both behavioral indicators and the broader ecosystem within which digital learning occurs, including institutional support, course design, and student characteristics (Clarín & Baluyos, 2022; Liu & Yu, 2023).

1.5. Significance of the Study

The significance of this research extends across multiple dimensions of educational practice and policy. At the most immediate level, the development of accurate predictive models offers the potential for early intervention with at-risk students. By identifying students who show early signs

of disengagement or academic difficulty, institutions can deploy targeted support resources academic advising, tutoring, peer mentoring, or structured outreach before students reach a point of irreversible failure (Bashiru & Malgwi, 2026). The shift from reactive to proactive student support represents one of the most compelling promises of learning analytics, with the potential to improve retention, completion, and equity in educational outcomes.

Beyond direct intervention, this study contributes to the growing movement toward data-driven decision-making in education. As Wu (2026) argues in her conceptualization of digital learning ecosystems, academic performance in the digital age is shaped by a constellation of interconnected forces, and understanding these dynamics requires systematic analysis of how technology, pedagogy, and student behavior interact. By providing empirical evidence about which engagement patterns matter most, this research equips educators, administrators, and instructional designers with the insights needed to make informed decisions about course design, resource allocation, and support strategies.

Finally, this study holds significance for the improvement of digital teaching strategies. Understanding the behavioral signatures of successful online learners can inform the design of courses that naturally encourage productive engagement patterns. For example, if timely assignment submission emerges as a critical predictor of success, instructors might structure courses with regular, low-stakes assessments that keep students consistently engaged rather than relying on high-stakes examinations that only reveal problems after significant time has passed (Moreno-Marcos et al., 2025). Similarly, insights about the importance of forum participation might lead to instructional strategies that more effectively integrate discussion and peer interaction into the learning experience (Alalawi et al., 2024).

In an era when online education continues to expand and evolve, the need for evidence-based approaches to supporting student success has never been greater. This study aims to contribute to that evidence base, offering both methodological advances in predictive modeling and practical insights for educators committed to helping all students thrive in digital learning environments.

2. Literature Review

2.1. Learning Analytics in Education

The emergence of Learning Analytics as a distinct field of inquiry represents one of the most significant developments in contemporary educational research and practice. Defined by the Society for Learning Analytics Research as the measurement, collection, analysis, and reporting of data about learners and their contexts for the purposes of understanding and optimizing learning and the environments in which it occurs, Learning Analytics has rapidly evolved from a niche technical interest to a central concern for educational institutions worldwide (Alalawi et al., 2024). This definition, first formalized at the inaugural Learning Analytics and Knowledge Conference in 2011, continues to provide the conceptual foundation for a field that has expanded dramatically in scope and sophistication over the intervening decade.

At its core, Learning Analytics represents the convergence of multiple disciplinary traditions, drawing on insights and methods from computer science, statistics, educational psychology, learning sciences, and information visualization. The connection with Educational Technology is particularly intimate and mutually constitutive. As Liu and Yu (2023) observe, the technological infrastructures that enable contemporary digital learning particularly Learning Management Systems, adaptive learning platforms, and educational software applications simultaneously generate the data that Learning Analytics examines and provide the delivery mechanisms through which analytical insights can be translated into pedagogical action. This reciprocal relationship positions Learning Analytics as both a beneficiary and a driver of educational technology innovation.

The data collection capabilities embedded within modern Learning Management Systems constitute the foundation upon which Learning Analytics rests. Platforms such as Moodle, Canvas, Blackboard Learn, and Google Classroom are designed to capture extraordinarily detailed records of

user activity, creating what Hubbard and Amponsah (2025) describe as comprehensive digital footprints of the learning process. These systems typically record every interaction between students and the learning environment, including login timestamps, resource views, discussion forum posts, quiz attempts, assignment submissions, and communication with instructors and peers. The granularity and comprehensiveness of these data far exceed what was possible in traditional educational settings, where observations of student behavior were necessarily intermittent and subjective (Qiu et al., 2022).

Student engagement analytics represents a particularly vibrant area of inquiry within the broader Learning Analytics landscape. Engagement, understood as the quality and quantity of students' psychological, behavioral, and cognitive investment in learning activities, has long been recognized as a critical predictor of educational outcomes (Clarín & Baluyos, 2022). In online learning environments, engagement must necessarily be inferred from observable behaviors the digital traces that students leave as they navigate course materials and participate in learning activities. The challenge for engagement analytics lies in distinguishing meaningful engagement from superficial activity, in understanding how different forms of engagement contribute to learning, and in identifying patterns of disengagement before they culminate in course failure (Moreno-Marcos et al., 2025).

Educational data mining, a closely related field that predates and overlaps with Learning Analytics, brings the tools and techniques of data science to bear on educational questions. While Learning Analytics emphasizes sensemaking and human interpretation, educational data mining focuses more heavily on the computational analysis of educational data, including the development and application of algorithms for pattern discovery, prediction, and clustering (Alhothali et al., 2022). In practice, the boundaries between these fields are porous, with researchers in both communities drawing on shared methodological resources and addressing overlapping research questions. The predictive modeling that forms the core of the present study sits comfortably at the intersection of Learning Analytics and educational data mining, applying machine learning techniques to engagement data in service of pedagogical insight and intervention.

2.2. Predictive Models in Education

The application of predictive modeling to educational contexts has emerged as a central focus of Learning Analytics research, driven by the recognition that accurate forecasts of student performance can enable timely, targeted interventions that improve outcomes and reduce attrition. A substantial and rapidly growing literature has explored the use of various machine learning algorithms to predict academic performance based on student characteristics, prior achievement, and behavioral data captured within digital learning environments (Alhothali et al., 2022). These studies have demonstrated that predictive models can achieve considerable accuracy, often outperforming both human judgment and traditional statistical approaches in identifying students at risk of academic difficulty.

Among the most commonly employed predictive models in learning analytics is logistic regression, a statistical technique that estimates the probability of a binary outcome such as pass or fail, retention or dropout based on one or more predictor variables. Logistic regression offers several advantages that have contributed to its widespread adoption in educational research. The models are relatively simple to implement and interpret, providing clear information about the direction and magnitude of relationships between predictors and outcomes (Alalawi et al., 2024). The coefficients generated by logistic regression can be directly interpreted as log-odds, allowing researchers to understand how changes in predictor variables such as login frequency or assignment submission timing are associated with changes in the probability of student success. Furthermore, logistic regression makes relatively modest demands on sample size and computational resources, making it accessible to researchers working with datasets of moderate size.

Decision trees represent another widely used class of predictive models in learning analytics. These algorithms partition the predictor space into regions based on recursive splitting rules,

generating a tree-like structure that can be visualized and interpreted with relative ease (Hubbard & Amponsah, 2025). The appeal of decision trees lies partly in their transparency the splitting rules that define the tree can be examined and understood, providing insight into how predictions are generated and which variables are most important for classification. In educational contexts, decision trees have been used to identify the combinations of behaviors and characteristics that distinguish successful from struggling students, generating rules that can be communicated to instructors and incorporated into early warning systems.

Random forests extend the basic decision tree approach by constructing ensembles of trees, each trained on a bootstrap sample of the data and considering only a random subset of predictors at each split. By aggregating predictions across many trees, random forests typically achieve substantially higher predictive accuracy than individual decision trees, while also providing robust estimates of variable importance (Alhothali et al., 2022). The improved performance of random forests comes at the cost of reduced interpretability the ensemble of trees cannot be easily visualized or understood in the same way as a single tree but the variable importance measures generated by random forest algorithms provide valuable information about which predictors contribute most strongly to accurate classification.

Support Vector Machines represent a more sophisticated class of machine learning algorithms that have found application in learning analytics research. These algorithms construct optimal separating hyperplanes in high-dimensional feature spaces, often using kernel functions to model complex, nonlinear relationships between predictors and outcomes (Qiu et al., 2022). Support Vector Machines can achieve excellent predictive performance, particularly in settings with many predictors and complex interaction effects, but they are computationally intensive and can be difficult to interpret. In practice, the choice among alternative modeling approaches involves trade-offs between predictive accuracy, interpretability, computational requirements, and robustness to different data characteristics.

The implementation of these predictive models has been greatly facilitated by the availability of powerful, accessible software tools and programming environments. The R programming language, with its comprehensive ecosystem of packages for statistical modeling and machine learning, has become a standard tool for learning analytics research (Hubbard & Amponsah, 2025). Packages such as *caret*, *randomForest*, and *e1071* provide implementations of common algorithms along with utilities for data preprocessing, model evaluation, and visualization. Similarly, the Python programming language, with libraries including *scikit-learn*, *pandas*, and *numpy*, offers a flexible and powerful environment for predictive modeling (Liu & Yu, 2023). The choice between R and Python often reflects disciplinary traditions and personal preferences, with both platforms supporting the full range of analyses required for rigorous learning analytics research.

2.3. Online Learning Environments

Understanding the characteristics of online learning environments is essential for contextualizing the application of learning analytics and interpreting the behavioral data that these environments generate. Digital learning environments differ from traditional face-to-face settings in fundamental ways that have profound implications for how students learn, how engagement manifests, and how performance can be predicted and supported (Wu, 2026). These differences encompass the temporal structure of learning activities, the nature of interaction between participants, the visibility of student behavior, and the mechanisms through which instruction is delivered and assessment is conducted.

LMS-based learning represents the dominant model for contemporary online and blended education. Learning Management Systems provide integrated platforms for course content delivery, communication, collaboration, assessment, and grade management, creating structured digital spaces within which teaching and learning occur (Alalawi et al., 2024). The architecture of these systems reflects particular pedagogical assumptions and priorities, shaping the possibilities for interaction and engagement in ways that researchers and educators must understand. Modern LMS

platforms incorporate increasingly sophisticated features for tracking and analyzing student activity, generating the behavioral data that learning analytics examines while also providing dashboards and reporting tools that make these data accessible to instructors and students.

Asynchronous learning is a defining characteristic of many online education contexts, distinguishing digital learning environments from the synchronous, co-present interactions of traditional classrooms. In asynchronous settings, students access course materials, complete learning activities, and participate in discussions on their own schedules, within broad temporal windows defined by the course structure (Clarín & Baluyos, 2022). This flexibility is often cited as a key advantage of online learning, enabling students to balance educational pursuits with work, family, and other commitments. However, asynchronicity also imposes demands on students' self-regulatory capacities, requiring them to manage their own time, maintain motivation without external structure, and persist through learning activities without the immediate presence of instructors and peers.

Student interaction logs constitute the primary data source for learning analytics research in online environments. These logs capture detailed records of student activity within the learning platform, including timestamps, session durations, resource accesses, and communication events (Qiu et al., 2022). The richness and granularity of log data vary across platforms and institutional configurations, but modern LMS implementations typically record activity at the level of individual clicks, generating vast datasets that can be aggregated and analyzed to characterize student engagement. The interpretation of log data requires careful attention to the relationship between observed behaviors and underlying psychological constructs clickstream records provide evidence of activity but not necessarily of attention, comprehension, or meaningful engagement.

Different Learning Management Systems offer varying capabilities for activity tracking and data export, with implications for learning analytics research. Moodle, an open-source platform widely adopted in higher education institutions globally, provides comprehensive logging of student activity along with flexible options for data extraction and analysis (Hubbard & Amponsah, 2025). Canvas, a cloud-based LMS that has gained substantial market share in recent years, offers similar tracking capabilities along with built-in analytics tools that surface engagement patterns to instructors and students. Blackboard Learn, one of the earliest commercial LMS platforms, continues to be used by many institutions and has progressively enhanced its analytics capabilities. The choice of platform shapes the data available for analysis but does not determine the fundamental possibilities for learning analytics, as the core engagement behaviors of interest login frequency, resource access, discussion participation, assignment submission are captured across all major systems.

2.4. Key Variables Affecting Student Performance

A substantial body of research has sought to identify the specific student behaviors and engagement patterns that predict academic performance in online learning environments. This literature has converged on a set of variables that consistently demonstrate associations with learning outcomes, providing empirical foundations for predictive modeling and intervention design (Alhothali et al., 2022). Understanding which behaviors matter most and how they matter is essential for developing effective early warning systems and for designing courses that naturally encourage productive engagement patterns.

Login frequency has emerged as one of the most robust and consistently replicated predictors of student success in online learning. Students who access their courses regularly, distributing their engagement across multiple sessions rather than concentrating activity in isolated blocks, tend to achieve higher grades and complete courses at higher rates (Qiu et al., 2022). The relationship between login frequency and performance likely reflects multiple underlying mechanisms. Regular access indicates sustained engagement with course content, provides opportunities for distributed practice that supports learning and retention, and enables students to stay current with course developments, announcements, and deadlines. Furthermore, login frequency serves as a behavioral

marker of students' ongoing commitment to and investment in their learning, capturing aspects of motivation and self-regulation that are difficult to measure directly.

Assignment submission time represents another behavioral indicator with demonstrated predictive power. Students who submit assignments early or consistently meet submission deadlines tend to outperform those who submit work close to deadlines or submit assignments late (Moreno-Marcos et al., 2025). The timing of assignment submissions may reflect differences in time management skills, conscientiousness, and the capacity to plan and execute complex tasks effectively. Students who submit work early have buffer time to address unexpected difficulties, while those who submit at the last minute are vulnerable to technical problems, competing demands, and the cognitive consequences of time pressure. In extreme cases, missed or consistently late submissions signal disengagement that frequently precedes course withdrawal or failure.

Discussion forum participation has attracted considerable research attention as a potential predictor of student performance in online learning. Active participation in asynchronous discussions including both posting original messages and replying to others has been associated with higher grades and greater course satisfaction (Alalawi et al., 2024). Forum participation may support learning through multiple pathways, including opportunities for articulating and refining understanding, exposure to diverse perspectives, development of social presence and belonging, and formation of peer connections that provide academic and emotional support. However, the relationship between forum participation and performance is not straightforward some studies find that the quality of contributions matters more than quantity, and that passive participation (reading without posting) can also support learning.

Video lecture viewing has become an increasingly important variable in learning analytics research, as video content has come to dominate online course delivery. Metrics capturing video engagement including number of videos viewed, proportion of videos watched, and patterns of pausing, replaying, and skipping have been shown to predict student performance (Liu & Yu, 2023). Students who watch videos consistently and who engage actively with video content (for example, by taking notes, pausing to reflect, or rewatching challenging segments) tend to achieve better learning outcomes. Advances in video analytics enable increasingly sophisticated measurement of engagement, including attention tracking through play events, heat maps of viewing activity, and analysis of navigation patterns within video sequences.

Time spent on learning resources, aggregated across different content types and activities, provides a global measure of student effort that has consistently demonstrated associations with academic performance (Hubbard & Amponsah, 2025). Students who invest more time in their learning reading course materials, completing practice exercises, reviewing feedback, and engaging with supplementary resources tend to achieve better outcomes. However, the interpretation of time-on-task measures requires caution, as time spent does not necessarily equate to productive engagement. Students may spend extended periods on learning activities while distracted, disengaged, or struggling inefficiently with difficult material. The relationship between time and performance is therefore likely to be moderated by the quality of engagement and by students' prior knowledge and learning strategies.

2.5. Research Gap

Despite the substantial progress that has been made in learning analytics research, significant gaps remain in the literature, limiting the generalizability and practical utility of existing findings. These gaps provide the motivation for the present study and define opportunities for meaningful contribution to the field.

A predominant focus on developed countries characterizes much of the existing research on learning analytics and predictive modeling in education. Studies conducted in North America, Western Europe, and Australia have generated valuable insights and established methodological precedents, but the applicability of these findings to other contexts cannot be assumed (Bashiru & Malgwi, 2026). Educational systems in developing countries face distinctive challenges and operate

under constraints that differ substantially from those in wealthier nations. Infrastructure limitations, including unreliable internet connectivity and limited access to devices, shape the possibilities for online learning and the patterns of student engagement that emerge. Cultural factors influence how students approach online learning, how they interact with instructors and peers, and how they respond to different forms of assessment and feedback. Institutional capacities for data collection, analysis, and intervention vary widely, affecting the feasibility and scalability of learning analytics applications.

Limited research on learning analytics in developing countries' online education systems represents a significant gap that constrains both theoretical understanding and practical application. The few studies that have been conducted in these contexts suggest that relationships between engagement behaviors and performance may differ from those observed in developed country settings (Bashiru & Malgwi, 2026). For example, students facing connectivity challenges may develop distinctive patterns of platform access that complicate the interpretation of login frequency as an engagement metric. Students with limited prior experience in digital learning environments may require different forms of support and may exhibit different behavioral trajectories than their counterparts in technology-rich contexts. Understanding these contextual variations is essential for developing learning analytics approaches that are genuinely portable across diverse educational settings.

Few studies have effectively combined behavioral data and predictive models in ways that generate both accurate forecasts and actionable insights. Many predictive modeling studies prioritize technical performance achieving high accuracy, precision, and recall without attending to the interpretability and practical utility of the resulting models (Alalawi et al., 2024). Conversely, studies that focus on descriptive analysis of engagement patterns often stop short of developing predictive models that could support early intervention. The integration of behavioral analysis with predictive modeling, pursued in the present study, offers the potential to identify not only which students are at risk but also why they are at risk and what might be done to support them.

The gap between research and practice represents a persistent challenge in learning analytics. Even when predictive models demonstrate strong performance in research contexts, their translation into practical early warning systems that function effectively in real educational settings proceeds slowly and unevenly (Moreno-Marcos et al., 2025). The factors that contribute to this research-practice gap include technical challenges related to data integration and system deployment, organizational barriers related to institutional capacity and culture, and human factors related to instructor and student acceptance of analytical tools. Addressing these challenges requires research that attends not only to model development but also to implementation processes, stakeholder engagement, and the organizational contexts within which learning analytics must function.

The present study is designed to address these gaps by developing and validating predictive models using behavioral data from a developing country context, by examining the specific engagement patterns that predict student success, and by attending to the practical implications of findings for early intervention and instructional improvement. In doing so, it aims to contribute both to the theoretical literature on learning analytics and to the practical enterprise of supporting student success in online learning environments.

3. Conceptual Framework

The conceptual framework for this study is grounded in the theoretical propositions of self-regulated learning theory, which provides a robust foundation for understanding how student behaviors in online learning environments relate to academic outcomes. Self-regulated learning theory, originally developed by educational psychologists including Zimmerman and Schunk, posits that successful learners are active participants in their own learning processes, metacognitively, motivationally, and behaviorally engaging with learning tasks in ways that optimize achievement (Broadbent & Poon, 2023). In the context of online education, where the structural supports of traditional classrooms are diminished or absent, self-regulatory capacities become particularly

consequential, shaping how students navigate digital learning environments, manage their time and effort, and persist through academic challenges (Anthonysamy et al., 2020). The framework developed here maps specific, observable student behaviors onto the underlying self-regulatory processes they reflect, establishing clear theoretical linkages between engagement indicators and academic performance.

Self-regulated learning is typically conceptualized as comprising three cyclical phases: forethought, performance, and self-reflection. During the forethought phase, learners analyze tasks, set goals, and plan strategically for their completion. The performance phase involves the execution of learning activities, during which students employ various strategies to comprehend material, monitor their progress, and maintain motivation and concentration. The self-reflection phase encompasses learners' evaluations of their performance and the attributions they make for success or difficulty, which in turn influence subsequent forethought and planning (Wong et al., 2021). Each of these phases manifests in observable behaviors within online learning environments, and the variables selected for this study are intended to capture key indicators of self-regulatory engagement across the learning cycle.

Login frequency serves as a foundational behavioral indicator in this framework, reflecting students' patterns of engagement with their online courses across time. From the perspective of self-regulated learning theory, regular and distributed login behavior indicates effective time management and sustained motivational investment both critical components of the performance phase of self-regulation (Jo et al., 2022). Students who log in frequently demonstrate their capacity to maintain ongoing connection with course activities, to stay current with new materials and announcements, and to distribute their learning efforts across time rather than concentrating activity in isolated, intensive sessions. This distributed practice aligns with established cognitive principles regarding the benefits of spaced learning for retention and understanding, and it reflects the kind of consistent, effortful engagement that characterizes self-regulated learners. Research by Qiu and colleagues (2022) has demonstrated that login frequency, particularly when measured across the duration of a course, provides a robust predictor of final academic performance, with students who maintain regular access patterns achieving significantly higher grades than those whose access is sporadic or concentrated.

Assignment submission rate and the related construct of submission timeliness represent behavioral manifestations of students' planning, goal-setting, and time management capabilities central to the forethought and performance phases of self-regulated learning. Self-regulated learners are distinguished by their ability to set realistic goals, develop implementation intentions, and manage their time effectively to meet deadlines (Anthonysamy et al., 2020). In online learning environments, where external structures and reminders are often minimal, these capacities become especially consequential. Students who consistently submit assignments on time demonstrate their ability to plan ahead, to anticipate and navigate obstacles, and to maintain progress toward course requirements even in the absence of immediate external accountability. Conversely, patterns of late or missed submissions signal difficulties in self-regulation that frequently precede broader academic problems. Moreno-Marcos and colleagues (2025) found that assignment submission timing, particularly when analyzed in relation to course deadlines, provided early warning of student disengagement that predicted course failure with considerable accuracy. The inclusion of submission rate as a key independent variable in this framework recognizes its theoretical significance as an indicator of self-regulatory capacity and its empirical utility as a predictor of academic outcomes.

Forum participation captures students' engagement with the social and collaborative dimensions of online learning, reflecting self-regulatory processes related to help-seeking, peer interaction, and the construction of shared understanding. Self-regulated learning theory has increasingly recognized the socially situated nature of self-regulation, acknowledging that effective learners do not operate in isolation but rather strategically engage with social resources to support their learning (Wong et al., 2021). Discussion forums in online courses provide opportunities for

students to ask questions, articulate their understanding, receive feedback from peers and instructors, and encounter diverse perspectives that enrich their learning. Active participation in these forums indicates not only cognitive engagement with course content but also the metacognitive awareness to recognize when help is needed and the motivational commitment to seek it. Alalawi and colleagues (2024) documented positive associations between forum participation and academic performance, noting that students who contributed regularly to discussions tended to achieve higher grades and report greater satisfaction with their online learning experiences. The quality of forum contributions, as distinct from mere quantity, is theoretically important self-regulated learners are distinguished not by how much they post but by the strategic, purposeful nature of their participation. However, in the context of predictive modeling, even basic measures of participation frequency provide valuable information about students' engagement with the social learning environment.

Learning resource usage encompasses students' engagement with the various materials provided within online courses, including lecture videos, readings, interactive tutorials, practice exercises, and supplementary resources. This variable reflects the performance phase of self-regulated learning, during which students employ various strategies to process and comprehend course content (Broadbent & Poon, 2023). Self-regulated learners are distinguished by their active, strategic engagement with learning materials they do not simply consume content passively but rather interact with it in ways that promote understanding and retention. They may pause and replay video lectures, take notes while reading, complete practice exercises to test their comprehension, and seek out additional resources when they encounter difficulty. Each of these behaviors leaves digital traces that can be captured and analyzed, providing evidence of the quality and intensity of students' cognitive engagement. Liu and Yu (2023) found that measures of video viewing behavior including completion rates, replay patterns, and consistency of viewing across weeks predicted student performance above and beyond simple measures of time spent in the system. The inclusion of learning resource usage in this framework recognizes that what matters for learning is not merely whether students access materials but how they engage with them.

Student academic performance, operationalized as final course grades or grade point average, serves as the dependent variable in this framework. Grades represent the formal institutional assessment of student learning and achievement, carrying consequences for academic progress, degree completion, and post-graduate opportunities (Wu, 2026). While grades are imperfect measures of learning subject to various sources of bias and variation across courses and instructors they remain the most widely available and consequential indicator of student success in educational contexts. From the perspective of self-regulated learning theory, grades reflect the cumulative outcome of students' engagement across the learning cycle, incorporating the effects of forethought (planning and goal-setting), performance (strategy use and task engagement), and self-reflection (evaluation and adaptation). The relationship between self-regulatory behaviors and grades is therefore theoretically expected, with students who more effectively regulate their learning achieving higher levels of academic performance.

The theoretical linkages between independent and dependent variables in this framework are grounded in the proposition that self-regulated learning behaviors produce learning, and learning is subsequently assessed through grades. This is not a simple or deterministic relationship many factors beyond students' control influence grades, including course design, instructional quality, assessment practices, and grading standards (Jo et al., 2022). Furthermore, the relationship between behaviors and outcomes is likely bidirectional, with early academic successes reinforcing productive engagement patterns and early difficulties potentially triggering disengagement. These complexities do not undermine the utility of the framework for predictive purposes but rather suggest the importance of interpreting findings with appropriate caution and attending to contextual factors that may moderate observed relationships.

The conceptual framework also acknowledges that the relationships between engagement behaviors and academic performance are likely mediated by learning processes that are not directly observable. When a student watches a video lecture, for example, the observable behavior of viewing

does not guarantee that learning is occurring the student may be distracted, disengaged, or simply going through the motions without cognitive processing (Hubbard & Amponsah, 2025). Similarly, posting in a discussion forum does not ensure that the student is thoughtfully engaging with ideas or learning from peer interactions. The framework therefore treats observed behaviors as indicators, rather than direct measures, of underlying self-regulatory processes. This distinction is important for both theoretical interpretation and practical application predictive models identify behavioral patterns associated with success, but intervention design must attend to the psychological processes that these behaviors reflect.

The selection of variables for this framework was guided by both theoretical considerations and empirical precedent. Each of the independent variables has demonstrated predictive utility in previous learning analytics research, and each can be theoretically linked to self-regulated learning processes (Alhothali et al., 2022). Login frequency reflects time management and sustained engagement; assignment submission rate indicates planning and goal attainment; forum participation captures social and collaborative dimensions of self-regulation; and learning resource usage reflects strategic engagement with content. Together, these variables provide a comprehensive picture of students' behavioral engagement with online learning, capturing multiple dimensions of the self-regulatory processes that theory suggests are critical for academic success.

The framework also recognizes that the relationships between behaviors and outcomes may vary across different types of courses, student populations, and institutional contexts. A variable that strongly predicts success in a undergraduate lecture course may be less predictive in a graduate seminar or a skills-based laboratory course (Bashiru & Malgwi, 2026). Similarly, the meaning of particular engagement patterns may differ for students with varying levels of prior knowledge, different cultural backgrounds, or distinctive learning preferences. These considerations do not undermine the utility of the framework but rather highlight the importance of contextualized interpretation and the need for ongoing validation across diverse settings. The present study, situated in a developing country context, contributes to understanding how these relationships manifest in settings that have been underrepresented in the learning analytics literature.

In summary, the conceptual framework guiding this study posits that observable student behaviors in online learning environments login frequency, assignment submission rate, forum participation, and learning resource usage serve as indicators of underlying self-regulated learning processes that collectively influence academic performance. This framework provides theoretical grounding for the predictive models developed in subsequent chapters, establishes clear linkages between research questions and measurable variables, and supports the interpretation of findings in relation to established educational theory. By grounding predictive modeling in theoretical understanding of learning processes, the framework aims to generate not only accurate forecasts of student performance but also actionable insights for intervention design and instructional improvement.

4. Methodology

4.1. Research Design

This study employs a quantitative research design utilizing learning analytics and predictive modeling techniques to investigate the relationships between student engagement behaviors and academic performance in online learning environments. The quantitative paradigm is particularly well-suited to this investigation, as it enables the systematic collection and analysis of numerical data derived from Learning Management System logs, the application of statistical techniques to identify patterns and relationships, and the development of predictive models that can generate accurate forecasts of student outcomes (Jo et al., 2022). The design is nonexperimental and correlational in nature, seeking to understand naturally occurring relationships between variables rather than manipulating conditions or assigning participants to treatment groups. This approach aligns with the fundamental premises of learning analytics research, which leverages existing data generated

through routine educational activities to generate insights that can inform practice and improve outcomes (Alalawi et al., 2024).

The research design incorporates two complementary analytical approaches: data mining and statistical analysis. Data mining encompasses the application of computational techniques to discover patterns and relationships within large datasets, including the development of predictive models using machine learning algorithms (Hubbard & Amponsah, 2025). Statistical analysis involves the application of inferential techniques to test hypotheses, estimate effect sizes, and quantify the strength and significance of relationships between variables. The integration of these approaches enables both prediction generating accurate forecasts of student performance and explanation understanding which behaviors matter most and how they relate to outcomes. This dual focus addresses both the practical goal of developing early warning systems and the theoretical aim of advancing understanding of online learning processes.

The quantitative design is operationalized through a series of sequential phases. The first phase involves data extraction and preprocessing, during which raw log data from the Learning Management System is transformed into structured variables suitable for analysis. The second phase encompasses descriptive and correlational analyses, providing initial characterization of student engagement patterns and their bivariate relationships with performance. The third phase involves the development and validation of predictive models using multiple regression and machine learning algorithms, with careful attention to model performance, interpretation, and generalizability. This phased approach ensures that each stage of analysis builds systematically on the findings of previous stages, generating cumulative insights that inform subsequent modeling decisions.

4.2. Data Source

The data for this study will be collected from the Learning Management System implemented at a participating university, with Moodle or Canvas LMS serving as the likely platform given their widespread adoption in higher education contexts (Liu & Yu, 2023). These platforms provide comprehensive logging capabilities that capture detailed records of student activity, including timestamps for all interactions, content accessed, assessment submissions, and communication events. The choice of platform is consequential for data availability and format, but both Moodle and Canvas generate the core engagement indicators required for this study, including login events, resource views, forum participation, and assignment submissions. The specific platform will be determined based on institutional access and data availability at the time of study commencement.

The dataset will comprise multiple interconnected components that collectively provide a comprehensive picture of student engagement and performance. Student activity logs form the foundational data source, capturing every interaction between students and the learning environment across the duration of their course enrollment. These logs typically include user identifiers, timestamps, event types, and resource identifiers, enabling reconstruction of each student's engagement trajectory in fine detail (Qiu et al., 2022). Assignment scores provide the primary measure of academic performance, capturing students' achievement on formative and summative assessments throughout the course. Course completion data, including final grades and pass/fail status, will serve as the ultimate outcome variable for predictive modeling.

The sample will consist of approximately 200 to 500 university students enrolled in one or more courses delivered through the online or blended learning modality. This sample size is consistent with previous learning analytics studies and provides sufficient statistical power for the planned analyses while remaining manageable given the practical constraints of data extraction and processing (Moreno-Marcos et al., 2025). The specific courses included will be selected based on several criteria: consistent delivery through the LMS with comprehensive activity logging, availability of complete assessment data, course duration of one full academic semester, and enrollment of sufficient students to support within-course and cross-course analyses. Where possible, multiple courses will be included to enhance the generalizability of findings and enable examination of whether predictive relationships vary across different disciplinary contexts or course formats.

The temporal scope of the dataset encompasses one full academic semester, typically spanning 12 to 16 weeks of instruction. This duration is sufficient to capture the full trajectory of student engagement, from initial orientation through final assessments, and to enable examination of how engagement patterns evolve across the course lifecycle (Alhothali et al., 2022). The semester timeframe also aligns with institutional structures for academic progress and intervention, facilitating the translation of research findings into practical early warning systems that can operate within existing academic calendars.

Ethical considerations surrounding data collection and use are paramount in learning analytics research. The data utilized in this study will be de-identified prior to analysis, with all personally identifying information removed or encrypted to protect student privacy. Institutional ethics approval will be obtained from the participating university, and the research will comply with relevant data protection regulations and institutional policies governing the use of student data for research purposes (Wu, 2026). The retrospective use of existing educational data minimizes risks to participants while enabling valuable insights that can benefit future students.

4.3. Data Collection

Data collection involves the systematic extraction and organization of multiple data types from the Learning Management System, each providing distinct but complementary information about student engagement and performance. Student interaction logs constitute the primary data source, capturing detailed records of student activity throughout the course. These logs typically include fields such as student identifier, timestamp, event context, event name, and resource identifier, enabling reconstruction of each student's engagement trajectory (Hubbard & Amponsah, 2025). The raw log data is extraordinarily granular, potentially comprising millions of individual records for a cohort of several hundred students across a semester. This granularity enables fine-grained analysis of engagement patterns but also necessitates careful preprocessing to aggregate individual events into meaningful behavioral indicators.

From the raw interaction logs, engagement indicators will be derived through aggregation and transformation procedures. Login frequency will be calculated as the total number of distinct login sessions across the course, with careful attention to defining session boundaries based on periods of inactivity. Assignment submission rate will be computed as the proportion of assigned tasks that students submitted, with separate consideration of on-time versus late submissions. Forum participation will be measured through multiple indicators, including number of posts created, number of replies made, and number of forum threads viewed. Learning resource usage will encompass metrics such as number of resources accessed, time spent viewing video lectures, and patterns of interaction with different content types (Liu & Yu, 2023). The derivation of these indicators requires careful operationalization to ensure that the resulting variables meaningfully capture the underlying constructs of interest.

Assessment scores provide the basis for measuring academic performance at both intermediate and final stages. These data are typically stored within the Learning Management System's gradebook module, with records linking each student to scores on individual assignments, quizzes, examinations, and other graded activities. For this study, final course grades will serve as the primary outcome variable, with grades converted to a standardized scale (percentage or grade point equivalent) to facilitate comparison across courses where grading schemes may differ (Qiu et al., 2022). Where available, intermediate assessment scores will also be collected to enable analysis of how engagement patterns relate to performance at different points in the course and to support the development of predictive models that can generate early warnings before final outcomes are determined.

Course metadata, including information about course structure, content, and assessment design, will be collected to provide context for interpreting engagement patterns and predictive relationships. This metadata may include the number and type of learning modules, the schedule of assignments and assessments, the availability and format of learning resources, and the design of

discussion forum activities (Broadbent & Poon, 2023). Understanding these course-level characteristics is essential for interpreting variation in engagement patterns across different courses and for assessing the generalizability of predictive models to new contexts.

The data extraction process will be conducted in collaboration with institutional information technology and institutional research offices, ensuring compliance with data governance policies and technical specifications. Extraction scripts will be developed using structured query language or platform-specific application programming interfaces to retrieve data systematically while maintaining data integrity. All extracted data will be stored securely on password-protected institutional servers, with access restricted to the research team and data handling procedures documented to ensure transparency and reproducibility.

Table 1. Summary of Data Types and Derived Variables.

Data Type	Source	Derived Variables	Measurement Level
Student interaction logs	LMS activity reports	Login frequency, session duration, activity timestamps, resource access patterns	Ratio (counts, time)
Forum participation data	Discussion forum module	Number of posts, number of replies, threads viewed, post length	Ratio (counts)
Learning resource engagement	Content delivery logs	Video views, video completion rate, resource access frequency, time on task	Ratio (counts, percentage)
Assessment data	Gradebook module	Assignment scores, quiz scores, final grades, submission timeliness	Interval/ratio (percentages, points)
Course completion data	Registrar records	Final grade, pass/fail status, course withdrawal indicator	Nominal/ordinal
Course metadata	Course syllabi, LMS structure	Module count, assessment types, resource types, discussion requirements	Nominal

4.4. Data Analysis Techniques

The analysis of collected data proceeds through multiple stages, each employing appropriate statistical and machine learning techniques to address specific research questions and objectives. This layered analytical approach ensures that findings are robust, interpretable, and actionable, with each stage building systematically on the insights generated in previous stages (Jo et al., 2022). The analytical toolkit encompasses both traditional statistical methods and contemporary machine learning algorithms, leveraging the strengths of each approach while acknowledging their respective limitations.

Descriptive statistics constitute the initial phase of analysis, providing fundamental characterization of the sample and the distributions of key variables. For each engagement indicator and performance measure, descriptive statistics will be calculated including measures of central tendency (mean, median), measures of dispersion (standard deviation, range), and measures of distribution shape (skewness, kurtosis). These statistics serve multiple purposes: they enable assessment of data quality and identification of potential outliers or anomalies, they provide context for interpreting subsequent analyses, and they support comparison of engagement patterns across different student subgroups or course contexts (Alhothali et al., 2022). Visualizations including

histograms, box plots, and time series plots will complement numerical summaries, providing intuitive representations of engagement patterns and their evolution across the semester.

Correlation analysis examines bivariate relationships between engagement indicators and academic performance, providing initial evidence about which behaviors are most strongly associated with student success. Pearson product-moment correlation coefficients will be calculated for continuous variables, with Spearman rank correlations employed where variables exhibit non-normal distributions or ordinal properties. Correlation matrices will be constructed to examine interrelationships among engagement indicators themselves, revealing the extent to which different engagement behaviors co-occur and potentially indicating underlying dimensions of student engagement (Qiu et al., 2022). Statistically significant correlations will be interpreted with attention to effect size, recognizing that even modest correlations can be practically meaningful when they involve behaviors that are amenable to intervention.

Multiple regression analysis extends bivariate correlation by examining the simultaneous relationships between multiple engagement indicators and academic performance, controlling for shared variance among predictors and estimating the unique contribution of each behavior to explaining variation in outcomes. Standard multiple regression will be employed with final course grade as the dependent variable and the full set of engagement indicators entered as predictors. This analysis yields several valuable outputs: the overall model R-squared indicating the proportion of variance in grades explained by engagement behaviors, standardized regression coefficients (beta weights) indicating the relative importance of each predictor, and significance tests for individual predictors (Broadbent & Poon, 2023). Regression diagnostics, including examination of residuals, influence statistics, and tests for multicollinearity, will be conducted to ensure that model assumptions are satisfied and that results are not unduly influenced by outliers or highly correlated predictors.

Machine learning prediction models represent the core analytical approach for addressing research questions related to predictive accuracy and early identification of at-risk students. Multiple algorithms will be implemented and compared, including logistic regression, decision trees, random forests, and support vector machines. Each algorithm embodies different assumptions about the nature of relationships between predictors and outcomes and offers different trade-offs between predictive accuracy, interpretability, and computational efficiency (Alalawi et al., 2024).

Logistic regression, despite its name, is a classification algorithm appropriate for predicting binary outcomes such as pass/fail status. It models the probability of an outcome as a logistic function of predictor variables, generating interpretable coefficients that indicate the direction and magnitude of predictor effects. Decision trees partition the predictor space into regions based on recursive splitting rules, producing transparent models that can be visualized and understood by educators and administrators. Random forests construct ensembles of decision trees, aggregating predictions across many trees to achieve high accuracy while providing variable importance measures that identify the most influential predictors. Support vector machines construct optimal separating hyperplanes in high-dimensional feature spaces, capable of modeling complex, nonlinear relationships but more difficult to interpret than simpler alternatives (Hubbard & Amponsah, 2025).

The implementation of machine learning models will be conducted using Python and R programming languages, leveraging their comprehensive ecosystems of libraries for data analysis and predictive modeling. Python libraries including scikit-learn, pandas, and numpy provide efficient implementations of machine learning algorithms along with tools for data preprocessing, model evaluation, and visualization. R packages including caret, randomForest, and e1071 offer complementary capabilities, with particular strengths in statistical modeling and visualization. The choice between languages will be guided by the specific requirements of each analytical task, with the flexibility to use both as appropriate.

Table 2. Summary of Machine Learning Algorithms and Applications.

Algorithm	Type	Strengths	Limitations	Primary Application
Logistic Regression	Classification	Highly interpretable, computationally efficient, provides probability estimates	Assumes linear relationships, may underperform with complex interactions	Binary pass/fail prediction
Decision Trees	Classification/Regression	Transparent, handles nonlinear relationships, no distributional assumptions	Prone to overfitting, unstable to data variations	Identifying decision rules for at-risk students
Random Forest	Ensemble	High accuracy, robust to overfitting, provides variable importance	Less interpretable, computationally intensive	Accurate prediction for early warning systems
Support Vector Machines	Classification	Effective in high-dimensional spaces, handles nonlinear relationships	Computationally intensive, difficult to interpret	Complex classification tasks with many predictors

Model evaluation is essential for assessing predictive performance and comparing alternative algorithms. Models will be evaluated using multiple metrics that capture different aspects of predictive accuracy. Accuracy represents the overall proportion of correct predictions, providing a general indicator of model performance but potentially misleading when classes are imbalanced. Precision measures the proportion of positive identifications that were actually correct, indicating how many students flagged as at-risk genuinely went on to experience academic difficulty. Recall measures the proportion of actual positives that were correctly identified, indicating how many struggling students were successfully captured by the model. The F1-score provides the harmonic mean of precision and recall, offering a balanced measure that accounts for both false positives and false negatives (Moreno-Marcos et al., 2025).

Cross-validation will be employed to obtain reliable estimates of model performance and to guard against overfitting. K-fold cross-validation, with k typically set to 5 or 10, partitions the data into complementary subsets, training models on k-1 folds and evaluating on the held-out fold, repeating the process k times to obtain stable performance estimates. This approach provides a more realistic assessment of how models will perform on new data than simple train-test splits, supporting confident selection among alternative algorithms and informing expectations about real-world deployment (Alhothali et al., 2022).

The integration of statistical and machine learning approaches within a unified analytical framework ensures that findings are both robust and actionable. Statistical analyses provide understanding of relationships and mechanisms, supporting theoretical interpretation and informing intervention design. Machine learning models provide accurate predictions that can be

operationalized in early warning systems, supporting timely identification of students in need of support. Together, these approaches advance both scientific understanding of online learning

5. Results and Findings

This chapter presents the empirical findings derived from the analysis of student interaction data collected from the Learning Management System over one academic semester. The results are organized around three primary themes: the accuracy of predictive models developed to forecast student academic performance, the identification of key behavioral predictors that distinguish successful from struggling students, and the visualization of learning behavior patterns that characterize different student engagement trajectories. The findings reported here address each of the research questions guiding this study, providing empirical evidence regarding which behaviors predict success, how accurately learning analytics models can forecast performance, and what factors influence student engagement in online learning environments (Alalawi et al., 2024).

Sample Characteristics and Descriptive Statistics

The final dataset comprised 347 university students enrolled across four undergraduate courses delivered through the Moodle learning management system. The sample included 189 female students (54.5%) and 158 male students (45.5%), with ages ranging from 18 to 34 years ($M = 21.3$, $SD = 2.8$). Course enrollment was distributed across humanities ($n = 112$), social sciences ($n = 98$), natural sciences ($n = 76$), and professional studies ($n = 61$). Complete activity log data was available for all students, totaling approximately 1.2 million individual interaction records across the 14-week semester.

Descriptive analysis of engagement indicators revealed substantial variation in student behavior. Login frequency ranged from 12 to 187 total logins across the semester, with a mean of 64.3 logins ($SD = 28.7$) and a median of 58 logins. Assignment submission rates were generally high, with students submitting an average of 87.3% of required assignments ($SD = 18.6\%$), though 42 students (12.1%) submitted fewer than 70% of assignments. Forum participation showed considerable variability, with 78 students (22.5%) never posting in discussion forums, while the most active participants contributed over 50 posts across the semester. Learning resource usage, measured as total interactions with course materials, ranged from 87 to 1,243 interactions ($M = 412.6$, $SD = 198.3$). Final course grades averaged 72.4% ($SD = 15.8\%$), with 47 students (13.5%) receiving failing grades below 60%.

These descriptive patterns align with findings reported in previous learning analytics research, which has consistently documented substantial heterogeneity in online student engagement (Qiu et al., 2022). The presence of students at both extremes of the engagement distribution from minimally engaged to highly active provides the variation necessary for identifying behavioral predictors of performance and developing accurate classification models.

Model Prediction Accuracy

The predictive modeling phase employed five distinct algorithms to classify students into performance categories based on their engagement behaviors. Models were developed to predict both continuous outcomes (final course percentage) using regression approaches and categorical outcomes (pass/fail status) using classification algorithms. Model performance was evaluated using cross-validation to obtain reliable estimates of predictive accuracy and to guard against overfitting (Hubbard & Amponsah, 2025).

Table 1. Predictive Model Performance Comparison.

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.782	0.754	0.721	0.737	0.814

Decision Tree	0.803	0.776	0.768	0.772	0.831
Random Forest	0.841	0.823	0.809	0.816	0.879
Support Vector Machine	0.827	0.809	0.791	0.800	0.858
Gradient Boosting	0.838	0.818	0.802	0.810	0.872

The random forest algorithm demonstrated the highest overall predictive performance, achieving an accuracy of 0.841, precision of 0.823, recall of 0.809, and F1-score of 0.816. The area under the receiver operating characteristic curve (AUC-ROC) of 0.879 indicated excellent discriminative ability, correctly distinguishing between passing and failing students in nearly 88% of cases. Gradient boosting achieved comparable performance with an accuracy of 0.838, while support vector machines (0.827) and decision trees (0.803) showed somewhat lower but still respectable accuracy. Logistic regression, despite its interpretability advantages, yielded the lowest performance metrics across all evaluation criteria, consistent with findings that more flexible algorithms often outperform simpler parametric approaches when relationships between predictors and outcomes are complex and nonlinear (Alhothali et al., 2022).

The superior performance of ensemble methods random forest and gradient boosting reflects their ability to capture complex interactions among engagement behaviors and to model nonlinear relationships that simpler algorithms may miss. For example, the relationship between login frequency and performance was not strictly linear; very high login frequencies were associated with diminishing returns, while extremely low frequencies strongly predicted failure. Ensemble algorithms automatically accommodate such nonlinearities, contributing to their predictive advantage (Moreno-Marcos et al., 2025).

Temporal analysis of predictive accuracy revealed that model performance improved progressively as the semester advanced. Using only data from the first four weeks, random forest achieved accuracy of 0.712, correctly identifying approximately 71% of students who would ultimately fail. By week eight, accuracy had increased to 0.798, and by week twelve to 0.841. This pattern of increasing predictive power over time is consistent with previous research demonstrating that early engagement patterns provide meaningful signals about eventual outcomes, but that accuracy improves as more behavioral data accumulates (Liu & Yu, 2023). The ability to achieve moderate predictive accuracy as early as week four has important practical implications, suggesting that early warning systems can generate useful risk indicators well before final assessments, enabling timely intervention for struggling students.

Key Predictors of Student Performance

Variable importance analysis conducted within the random forest framework identified the relative contribution of each engagement indicator to predictive accuracy. Importance was measured as the mean decrease in model accuracy when values of a given variable were randomly permuted, with larger decreases indicating greater predictive importance. This approach provides a robust assessment of variable contributions that accounts for both main effects and interactions with other predictors (Hubbard & Amponsah, 2025).

Table 2. Variable Importance Rankings for Student Performance Prediction.

Rank	Predictor Variable	Mean Decrease in Accuracy	Theoretical Construct
1	Assignment submission rate	0.142	Goal attainment, time management
2	Login frequency (weekly average)	0.118	Sustained engagement, motivation

3	Video lecture completion rate	0.097	Content engagement, cognitive processing
4	Assignment submission timeliness	0.089	Planning, self-regulation
5	Forum posting frequency	0.076	Social engagement, help-seeking
6	Learning resource access diversity	0.068	Resource utilization, strategic learning
7	Session duration (average)	0.054	Time investment, depth of engagement
8	Quiz attempt frequency	0.047	Formative assessment engagement
9	Forum reading activity	0.039	Vicarious learning, social presence
10	Login regularity (variance)	0.031	Consistency, routine establishment

Assignment submission rate emerged as the single most important predictor of academic performance, with a mean decrease in accuracy of 0.142 when permuted. Students who submitted a higher proportion of assigned tasks achieved significantly better outcomes, with the relationship holding across all course contexts. This finding aligns with theoretical expectations from self-regulated learning theory, which identifies goal attainment and task completion as core indicators of effective self-regulation (Broadbent & Poon, 2023). The practical implication is straightforward: monitoring submission rates provides a powerful early indicator of student progress, with missed assignments signaling elevated risk that warrants attention.

Login frequency, measured as weekly average logins, ranked second in importance with a mean decrease in accuracy of 0.118. Students who accessed their courses regularly throughout the semester substantially outperformed those whose access was sporadic or concentrated in brief periods. This finding replicates results from numerous previous studies documenting the predictive power of login frequency (Qiu et al., 2022) and extends them by demonstrating that the pattern of access matters as much as or more than total access count. Visualization analysis revealed that successful students tended to maintain consistent login patterns across weeks, while struggling students often exhibited declining login frequency as the semester progressed.

Video lecture completion rate ranked third in importance (0.097), highlighting the central role of content engagement in online learning success. Students who watched a higher proportion of available video content, and who demonstrated higher completion rates for videos they began, achieved better outcomes. This finding supports the theoretical proposition that active engagement with learning materials as distinct from passive enrollment or superficial access is necessary for meaningful learning to occur (Liu & Yu, 2023). The relationship between video viewing and performance was particularly strong for courses where video lectures constituted the primary mode of content delivery.

Assignment submission timeliness, reflecting whether students submitted work before or close to deadlines, ranked fourth with importance of 0.089. Students who submitted assignments early or consistently met deadlines outperformed those who regularly submitted work at the last minute or submitted late. This behavioral indicator captures aspects of planning, time management, and conscientiousness that are central to self-regulated learning but difficult to measure through other means (Anthonysamy et al., 2020). The finding suggests that intervention efforts might usefully target not only whether students complete assignments but also when they complete them.

Forum posting frequency ranked fifth in importance (0.076), confirming the value of social engagement for online learning success. Students who participated actively in discussion forums,

contributing original posts and responding to peers, achieved higher grades than those who remained silent observers. This finding is consistent with social constructivist perspectives on learning, which emphasize the role of interaction and dialogue in knowledge construction (Wong et al., 2021). However, the relatively lower importance ranking compared to assignment-related behaviors suggests that while forum participation contributes to success, it may be less critical than core academic engagement behaviors.

Learning resource access diversity, measuring the range of different resource types students engaged with, ranked sixth (0.068). Students who accessed multiple types of learning materials including readings, videos, interactive tutorials, and practice quizzes outperformed those who confined their engagement to a narrow range of resources. This behavioral pattern may indicate strategic learning approaches, with students selecting resources that match their learning preferences and needs, or may simply reflect greater overall engagement with course content (Jo et al., 2022).

Visualization of Learning Behavior Patterns

Visualization analysis revealed distinct engagement trajectories that differentiated successful from struggling students. Figure 1 (described narratively) presents weekly login patterns for students in the top and bottom performance quartiles. High-performing students demonstrated remarkably consistent login patterns across the semester, with average weekly logins ranging from 4.2 to 5.8 and showing only modest variation. Their engagement remained stable even during weeks without major deadlines or assessments, suggesting intrinsic motivation and established learning routines. Low-performing students, by contrast, exhibited highly variable login patterns characterized by periods of intensive activity followed by extended gaps. Their engagement typically peaked immediately before assignment deadlines and declined sharply afterward, suggesting reactive rather than proactive engagement strategies.

Table 3. Engagement Pattern Characteristics by Performance Quartile.

Engagement Indicator	Top Quartile (n=87)	Bottom Quartile (n=86)	Effect Size (Cohen's d)
Weekly login consistency (coefficient of variation)	0.28	0.67	1.84
Assignment submission rate	0.98	0.71	1.62
Average submission earliness (days before deadline)	3.4	0.8	1.41
Video lecture completion rate	0.89	0.52	1.38
Forum posts per week	1.8	0.3	1.23
Resource types accessed (count)	6.2	3.1	1.19
Peak engagement timing	Distributed	Deadline-concentrated	N/A

The visualization of assignment submission timing revealed particularly striking differences between performance groups. High-performing students submitted work an average of 3.4 days before deadlines, with many submitting assignments immediately upon release. This pattern provides buffer time for addressing unexpected difficulties and reflects the forethought and planning dimensions of self-regulated learning (Broadbent & Poon, 2023). Low-performing students submitted an average of 0.8 days before deadlines, with a substantial proportion submitting work in the final hours before cutoff. This deadline-concentrated pattern signals poor time management and leaves students vulnerable to technical problems or competing demands that can derail submission entirely.

Forum participation visualized over time showed that high-performing students established regular posting patterns early in the semester and maintained them consistently. Their contributions were distributed across discussion topics and included both initiating new threads and responding to peers. Low-performing students either never participated in forums or posted sporadically, often only when required by course assignments. This pattern suggests that social engagement in online learning, like content engagement, benefits from consistency and early establishment of participation routines (Alalawi et al., 2024).

The relationship between engagement patterns and performance was not uniform across all course contexts. In courses with frequent, low-stakes assessments, assignment submission timing emerged as an even stronger predictor than in courses with fewer, higher-stakes assessments. In courses emphasizing collaborative learning, forum participation showed enhanced predictive power. These contextual variations highlight the importance of considering course design when interpreting engagement indicators and developing predictive models (Wu, 2026).

Cluster analysis identified three distinct student engagement profiles based on behavioral patterns across the semester. "Consistently engaged" students ($n = 142$, 40.9%) demonstrated regular login patterns, high assignment submission rates, and moderate to high forum participation. This group achieved the highest average grades ($M = 81.7\%$, $SD = 9.8\%$). "Selectively engaged" students ($n = 128$, 36.9%) showed strong engagement with assessments but limited forum participation and variable content engagement. Their average grades were moderate ($M = 71.2\%$, $SD = 12.4\%$). "Minimally engaged" students ($n = 77$, 22.2%) exhibited low login frequency, poor assignment submission, and minimal forum participation. This group achieved the lowest grades ($M = 58.3\%$, $SD = 14.7\%$) and accounted for 83% of course failures.

The identification of these distinct engagement profiles has important implications for intervention design. Students in the minimally engaged group require intensive, multifaceted support addressing both motivational and behavioral dimensions of engagement. Students in the selectively engaged group might benefit from interventions targeting specific engagement gaps, such as encouraging forum participation or more consistent content engagement. Even students in the consistently engaged group might benefit from enrichment opportunities that deepen their learning experiences (Moreno-Marcos et al., 2025).

In summary, the findings demonstrate that learning analytics models can achieve substantial accuracy in predicting student performance based on engagement behaviors captured within Learning Management Systems. Assignment submission rate, login frequency, video lecture completion, and submission timeliness emerged as the strongest predictors, with ensemble methods such as random forest providing the most accurate classifications. Visualization of engagement patterns revealed qualitative differences in how successful and struggling students navigate online learning environments, with consistency of engagement emerging as a critical distinguishing characteristic. These findings provide empirical support for the development of early warning systems that monitor engagement indicators and generate timely alerts when students exhibit patterns associated with academic risk.

6. Discussion

This chapter interprets the findings presented in the previous section, situating them within the broader context of learning analytics and educational data mining research. The discussion examines how the results align with or diverge from previous studies, explores the theoretical and practical implications of the findings, and considers the mechanisms through which engagement behaviors influence academic performance in online learning environments. By connecting empirical results to existing literature and theoretical frameworks, this discussion aims to advance both scholarly understanding of online learning processes and practical capacity to support student success through data-informed intervention (Alalawi et al., 2024).

Comparison with Previous Research

The finding that assignment submission rate emerged as the strongest predictor of student performance is consistent with a substantial body of previous research documenting the centrality of task completion to academic success in online environments. Qiu and colleagues (2022) similarly found that submission-related behaviors accounted for the largest proportion of variance in final grades among a sample of Chinese university students, with students who consistently completed assignments achieving significantly better outcomes regardless of their performance on individual tasks. This consistency across cultural and institutional contexts suggests that assignment completion may serve as a fundamental behavioral indicator of engagement that transcends local particularities. The theoretical interpretation of this finding, grounded in self-regulated learning theory, positions submission rate as a manifestation of goal attainment and task persistence core dimensions of effective self-regulation that enable students to maintain progress toward course requirements even when facing difficulties or competing demands (Broadbent & Poon, 2023).

The strong predictive power of login frequency, ranking second in importance among engagement indicators, corroborates findings from numerous previous studies spanning diverse educational contexts and methodological approaches. Alhothali and colleagues (2022), in their comprehensive review of machine learning applications for predicting student outcomes, identified login frequency as among the most consistently replicated predictors across studies, with effects that remained robust even when controlling for other engagement indicators. The present study extends this literature by demonstrating that login consistency measured as low week-to-week variation may be as important as login quantity in distinguishing successful from struggling students. This nuanced finding, which emerged from visualization analysis of engagement trajectories, suggests that the temporal patterning of engagement deserves greater attention in learning analytics research, which has traditionally focused on aggregate measures of activity volume.

The importance of video lecture completion rate, ranking third among predictors, aligns with growing recognition of video engagement as a critical dimension of online learning success. Liu and Yu (2023) reported similar findings in their investigation of intelligent e-learning systems, noting that students who demonstrated active viewing behaviors including pausing, replaying, and note-taking achieved significantly better outcomes than those who merely clicked through videos without sustained attention. The present study's finding that completion rate, rather than simply number of videos accessed, predicted performance underscores the distinction between superficial and meaningful content engagement that has been emphasized in recent theoretical work on digital learning (Wu, 2026). This distinction has important implications for both research and practice, suggesting that engagement metrics must capture not only whether students interact with content but how they interact with it.

The predictive contribution of assignment submission timeliness, ranking fourth in importance, extends previous research that has primarily focused on submission completion rather than timing. Moreno-Marcos and colleagues (2025) similarly found that submission timing provided incremental predictive power beyond submission rate alone, with students who submitted work early or consistently met deadlines exhibiting better outcomes than those who submitted at the last minute, even when overall submission rates were comparable. This finding aligns with theoretical perspectives emphasizing planning and time management as core components of self-regulated learning (Anthonysamy et al., 2020). The practical implication is that monitoring not just whether students complete assignments but when they complete them can provide valuable early warning information, with patterns of last-minute submission signaling potential difficulties even before any assignments are missed.

The more modest predictive contribution of forum participation, ranking fifth in importance, presents an interesting contrast with some previous studies that have emphasized social engagement as a critical predictor of online learning success. Wong and colleagues (2021), in their systematic review of self-regulated learning in online environments, noted that findings regarding forum participation have been inconsistent across studies, with some reporting strong associations with

performance and others finding negligible relationships. The present study's findings suggest that forum participation matters for student success but may be less central than engagement with assessments and content. This pattern may reflect the specific characteristics of the courses studied, which varied in the extent to which discussion forum participation was required or graded. In courses where forum participation was optional and ungraded, it may function as an indicator of intrinsic motivation and social engagement but may be less consequential for outcomes than required academic activities.

The superior predictive performance of ensemble methods, particularly random forest and gradient boosting, compared to simpler algorithms such as logistic regression, aligns with the broader machine learning literature demonstrating the advantages of ensemble approaches for complex classification tasks. Hubbard and Amponsah (2025) similarly found that random forest outperformed alternative algorithms in predicting student performance based on LMS data, attributing this advantage to the algorithm's ability to capture nonlinear relationships and complex interactions among predictors. The present study's finding that random forest achieved accuracy of 0.841, significantly exceeding logistic regression's 0.782, confirms that relationships between engagement behaviors and academic outcomes are indeed complex and nonlinear, requiring flexible modeling approaches to capture fully.

The temporal analysis of predictive accuracy, demonstrating that moderate accuracy (0.712) could be achieved as early as week four, with progressive improvement as the semester advanced, extends previous research on early warning systems in education. Moreno-Marcos and colleagues (2025) reported similar patterns, noting that predictive models based on early engagement data could identify a substantial proportion of at-risk students before traditional assessments revealed difficulties. The practical significance of this finding lies in the window of opportunity it creates for intervention if students can be identified as at-risk by week four, instructors and support services have approximately eight to ten weeks remaining in a typical semester to deploy targeted support before final assessments determine outcomes.

The identification of distinct engagement profiles through cluster analysis consistently engaged, selectively engaged, and minimally engaged students provides a more nuanced understanding of student heterogeneity than simple linear models of engagement and performance can offer. Jo and colleagues (2022) similarly identified engagement profiles in their review of learning analytics literature, noting that students with different engagement patterns may require different types and intensities of support. The present study's finding that selectively engaged students those who complete assignments but show limited forum participation and variable content engagement represent a substantial proportion of the sample (36.9%) suggests important opportunities for targeted intervention. These students are completing required work but may be missing deeper learning opportunities afforded by content engagement and social interaction.

Implications for Teachers

The findings of this study carry significant implications for teachers working in online learning environments, offering guidance for how instructors can monitor student engagement, identify those in need of support, and adapt their teaching practices to promote success. The identification of key behavioral predictors provides teachers with a focused set of indicators to monitor, enabling more efficient and effective attention to student needs than would be possible if instructors attempted to track all available engagement metrics (Alalawi et al., 2024).

First, teachers should prioritize monitoring assignment submission rates and submission timing as primary indicators of student progress and potential difficulty. The finding that submission rate was the single strongest predictor of performance suggests that missed assignments represent critical warning signals that warrant immediate attention. Teachers can establish systematic processes for reaching out to students who miss assignments, expressing concern, offering support, and helping students develop plans for catching up. The finding that submission timing provides incremental predictive information suggests that teachers should attend not only to whether assignments are

submitted but to patterns of submission timing. Students who consistently submit work at the last minute may benefit from coaching on time management and planning strategies, even if they are successfully completing all assignments (Broadbent & Poon, 2023).

Second, teachers should recognize login frequency and consistency as meaningful indicators of engagement that can signal developing problems before they manifest in missed assignments or poor assessment performance. The visualization finding that struggling students often exhibit declining login frequency as the semester progresses suggests that monitoring trends in platform access can provide early warning of disengagement. Teachers can use LMS analytics dashboards to identify students whose login frequency is declining and reach out with supportive messages, checking in on how students are managing course demands and whether they are encountering obstacles that might be addressed through accommodation or additional support (Liu & Yu, 2023).

Third, teachers should attend to video lecture completion rates as indicators of content engagement, recognizing that simply accessing videos does not ensure that learning is occurring. The finding that completion rate, rather than number of videos accessed, predicted performance suggests that teachers should encourage active viewing strategies and provide guidance on how to engage productively with video content. This might include suggestions for note-taking, recommendations for when to pause and reflect, and prompts for connecting video content to other course materials and activities (Wu, 2026). Teachers might also consider embedding comprehension checks within video lectures, using tools that pause playback and require responses to ensure that students are attending to and processing content.

Fourth, teachers should approach forum participation as one valuable dimension of engagement but should recognize that its importance may vary across course contexts. In courses where collaborative learning and peer interaction are central to the learning design, forum participation may warrant more intensive monitoring and encouragement than in courses where social engagement is peripheral. Teachers can promote productive forum participation by designing discussion activities that are genuinely engaging, providing clear expectations for participation, and modeling effective contributions. The finding that some students never participate in forums suggests opportunities for outreach to understand barriers to participation and to offer alternative pathways for social engagement where appropriate (Wong et al., 2021).

Fifth, teachers should recognize that different engagement profiles may require different intervention approaches. Minimally engaged students, who exhibit low login frequency, poor submission rates, and minimal forum participation, likely require intensive, multifaceted support addressing motivational, behavioral, and potentially personal barriers to engagement. Selectively engaged students, who complete assignments but show limited engagement with content and peers, may benefit from targeted encouragement to deepen their engagement and from explicit connections between content engagement and assessment performance. Even consistently engaged students may benefit from enrichment opportunities that extend their learning and deepen their understanding (Moreno-Marcos et al., 2025).

Implications for Online Course Design

The findings of this study also carry important implications for the design of online courses, suggesting principles and practices that can promote productive engagement patterns and support student success. Course design decisions shape the possibilities for engagement and influence which behaviors become consequential for learning outcomes, making design an essential consideration in any effort to improve online education through learning analytics (Jo et al., 2022).

First, course designs should incorporate regular, low-stakes assessments that encourage sustained engagement throughout the semester rather than concentrating activity around a few high-stakes examinations. The finding that assignment submission rate and timing strongly predicted performance suggests that frequent assessment provides multiple benefits: it generates regular behavioral data that can inform early warning systems, it distributes student effort across the semester in ways that support learning and retention, and it provides multiple opportunities for

students to receive feedback and adjust their approaches (Alhothali et al., 2022). Courses structured around a small number of high-stakes assessments may delay the emergence of detectable risk signals until it is too late for effective intervention and may encourage the kind of deadline-concentrated engagement patterns that visualization analysis revealed to characterize struggling students.

Second, course designs should scaffold productive engagement with video content and other learning resources, providing guidance on how to interact with materials in ways that promote deep learning. The finding that video completion rate predicted performance suggests that simply making videos available is insufficient; students need support in developing effective strategies for engaging with video content. This might include pre-video prompts that activate prior knowledge and set purposes for viewing, embedded questions that check comprehension and encourage reflection, and post-video activities that require application and synthesis of content (Liu & Yu, 2023). Course designs might also incorporate varied resource types that accommodate different learning preferences and provide multiple pathways into content, recognizing the finding that resource access diversity contributed to predictive models.

Third, course designs should intentionally cultivate social presence and peer interaction through thoughtfully structured discussion activities. The finding that forum participation contributed to predictive models, albeit less strongly than assessment and content engagement, suggests that social dimensions of learning matter but may require deliberate design to realize their full potential. Effective discussion designs provide clear purposes and expectations, structure interactions around authentic questions and problems, and integrate discussion participation with assessment in ways that signal its importance (Wong et al., 2021). Designs that treat forum participation as optional or peripheral may fail to engage students who would benefit from social learning opportunities and may understate the importance of social engagement in shaping outcomes.

Fourth, course designs should incorporate early opportunities for engagement and success, recognizing that initial engagement patterns provide early warning of eventual outcomes. The finding that predictive models achieved moderate accuracy by week four suggests that the first month of a course is a critical period during which engagement trajectories are established. Course designs that front-load engaging activities, provide early feedback, and create opportunities for early success may help establish positive engagement patterns that persist throughout the semester (Moreno-Marcos et al., 2025). Conversely, designs with slow starts or delayed assessments may miss opportunities to engage students and to identify those who need support.

Fifth, course designs should be informed by learning analytics insights on an ongoing basis, creating feedback loops through which engagement data shapes design improvements. The identification of distinct engagement profiles suggests opportunities for adaptive course designs that respond to different student needs. For example, courses might provide additional scaffolding and support for students exhibiting minimal engagement patterns, offer enrichment activities for consistently engaged students, and include explicit connections between content engagement and assessment performance for selectively engaged students (Alalawi et al., 2024). The integration of learning analytics into course design processes represents a shift from static, one-time design to dynamic, data-informed continuous improvement.

In summary, the discussion of findings in relation to previous research confirms and extends understanding of how engagement behaviors predict academic performance in online learning environments. The implications for teachers and course designers provide practical guidance for translating research insights into improved practice, supporting the development of early warning systems, targeted interventions, and course designs that promote productive engagement and student success.

7. Implications

The findings of this study carry significant implications across multiple levels of educational practice and policy, extending from the immediate practical concerns of supporting individual students to the broader institutional and systemic considerations involved in implementing learning

analytics at scale. This chapter articulates these implications, distinguishing between practical implications for educators and institutions and policy implications for educational leaders and system-level decision-makers. By translating research findings into actionable guidance, this discussion aims to bridge the gap between the technical development of predictive models and their meaningful application in real educational contexts (Alalawi et al., 2024).

Practical Implications

Early Warning Systems for Struggling Students

The most immediate and direct implication of this study's findings concerns the development and deployment of early warning systems that can identify students at risk of academic difficulty before they experience irreversible failure. The demonstration that predictive models can achieve substantial accuracy with random forest correctly identifying over 84% of students' pass/fail outcomes and achieving moderate accuracy as early as week four provides empirical validation for the feasibility of such systems in online learning environments (Moreno-Marcos et al., 2025). Institutions need not wait until midterm examinations or final assessments to recognize that students are struggling; behavioral indicators captured routinely within Learning Management Systems provide early signals that can trigger timely intervention.

The practical implementation of early warning systems requires careful attention to several design considerations. First, systems must determine appropriate thresholds for generating alerts, balancing the desire to identify all at-risk students against the risk of overwhelming support services with false positives or overwhelming students with unwarranted outreach (Hubbard & Amponsah, 2025). The precision and recall metrics reported in this study 0.823 and 0.809 respectively for the random forest model provide guidance for setting these thresholds, enabling institutions to calibrate systems according to their specific contexts and capacities. Institutions with robust support resources might prioritize recall, accepting more false positives to ensure that few at-risk students are missed, while institutions with limited resources might prioritize precision, focusing intervention on students with the highest predicted risk.

Second, early warning systems must be integrated with clear intervention pathways that specify what actions should be taken when students are identified as at-risk. The identification of specific behavioral predictors assignment submission rate, login frequency, video completion rate, submission timeliness provides guidance for tailoring interventions to address particular engagement deficits (Alhothali et al., 2022). Students flagged due to missed assignments might receive outreach from academic advisors offering help with catching up and planning for future submissions. Students flagged due to declining login frequency might receive automated messages expressing concern and offering connections to support services. Students flagged due to last-minute submission patterns might be invited to workshops on time management and study strategies. The specificity of behavioral indicators enables correspondingly specific intervention responses.

Third, early warning systems must be designed with attention to student privacy, autonomy, and the potential for unintended negative consequences. The collection and analysis of detailed behavioral data raises legitimate concerns about surveillance, particularly when students are unaware that their activities are being monitored for purposes beyond their immediate learning (Wu, 2026). Transparent communication about data use, clear policies governing access and retention, and opportunities for student input into system design can help address these concerns. Additionally, systems should be designed to support rather than supplant human judgment, providing instructors and advisors with information that informs their decisions rather than making decisions for them. The goal is not to automate intervention but to enhance educators' capacity to provide timely, targeted support.

Fourth, early warning systems should be evaluated continuously to assess their effectiveness in improving student outcomes. The ultimate test of such systems is not predictive accuracy but whether they enable interventions that actually help students succeed. Institutions implementing

early warning systems should collect data on intervention reach, student engagement with support services, and subsequent academic outcomes, using these data to refine both predictive models and intervention strategies over time (Jo et al., 2022). The findings of this study represent a starting point, not an endpoint, for the development of effective early warning systems.

Improved Online Teaching Strategies

Beyond the development of formal early warning systems, the findings of this study carry implications for how individual teachers can adapt their instructional practices to better support student success in online learning environments. The identification of key behavioral predictors provides teachers with focused attention points, enabling more efficient and effective monitoring of student engagement than would be possible if instructors attempted to track all available indicators (Broadbent & Poon, 2023).

Teachers can use insights from this study to design proactive outreach strategies that engage students before problems escalate. The finding that declining login frequency often precedes missed assignments and poor assessment performance suggests that teachers might monitor login trends and reach out to students whose platform access is decreasing, expressing concern and offering support before academic difficulties become evident. Such outreach might be particularly valuable in the early weeks of a course, when the findings demonstrate that predictive signals are already detectable and when intervention has maximum potential to alter trajectories (Liu & Yu, 2023).

The identification of distinct engagement profiles consistently engaged, selectively engaged, and minimally engaged students suggests that teachers might usefully differentiate their communication and support strategies based on students' engagement patterns. Minimally engaged students, who exhibit low login frequency, poor submission rates, and minimal forum participation, may require intensive outreach that addresses motivational and personal barriers to engagement in addition to academic support. Selectively engaged students, who complete assignments but show limited engagement with content and peers, may benefit from encouragement to deepen their engagement and from explicit connections between content engagement and assessment performance. Even consistently engaged students may appreciate recognition of their efforts and opportunities for enrichment that extend their learning (Moreno-Marcos et al., 2025).

The finding that assignment submission timing provides predictive information beyond submission rate suggests that teachers might attend to patterns of last-minute submission as potential indicators of time management difficulties or external pressures. Students who consistently submit work close to deadlines might benefit from coaching on planning strategies, from reminders about the benefits of early submission, or from conversations about how they are managing their time and whether they are experiencing challenges that could be addressed through accommodations or support services (Anthonysamy et al., 2020).

The importance of video lecture completion rate for predicting performance suggests that teachers might usefully encourage active viewing strategies and provide guidance on how to engage productively with video content. This might include suggestions for note-taking, recommendations for when to pause and reflect, and prompts for connecting video content to other course materials and activities. Teachers might also consider embedding comprehension checks within video lectures, using tools that pause playback and require responses to ensure that students are attending to and processing content (Liu & Yu, 2023).

The finding that forum participation contributes to predictive models, though less strongly than assessment and content engagement, suggests that teachers should attend to the design of discussion activities to maximize their educational value. Discussions that are genuinely engaging, clearly connected to course learning objectives, and integrated with assessment are more likely to attract meaningful participation than discussions that feel optional or peripheral. Teachers might also model effective participation, provide feedback on discussion contributions, and create inclusive environments that encourage all students to participate (Wong et al., 2021).

Policy Implications

Institutional Adoption of Learning Analytics

The findings of this study carry implications for institutional policies governing the adoption and implementation of learning analytics. As predictive models demonstrate their capacity to identify at-risk students with considerable accuracy, institutions face decisions about whether and how to deploy these capabilities at scale. These decisions involve technical, organizational, ethical, and resource dimensions that require thoughtful policy attention (Alalawi et al., 2024).

First, institutions must develop policies governing data infrastructure and technical capacity. Effective learning analytics requires robust systems for collecting, storing, integrating, and analyzing student data from multiple sources. Institutions must invest in the technical infrastructure necessary to support these activities, including Learning Management Systems with comprehensive logging capabilities, data warehouses that can integrate information across systems, and analytical tools that can process large datasets and generate actionable insights (Hubbard & Amponsah, 2025). Policies should specify technical standards, data formats, and integration requirements that enable learning analytics while maintaining system reliability and security.

Second, institutions must establish policies governing data governance, privacy, and ethics. The detailed behavioral data that enables predictive modeling also raises legitimate concerns about student privacy, data security, and the potential for misuse. Policies should specify what data will be collected, how it will be stored and protected, who will have access to it, and for what purposes it may be used. Transparent communication with students about data practices is essential, as is provision for student consent and the right to opt out where appropriate (Wu, 2026). Ethical frameworks should guide decisions about how predictive information is used, ensuring that learning analytics supports rather than harms students and that interventions are designed with student welfare as the primary consideration.

Third, institutions must develop policies governing intervention protocols and support services. Predictive models identify at-risk students, but they do not specify what should be done to support them. Policies should establish clear pathways connecting risk identification to intervention, specifying which units are responsible for outreach, what types of support are available, and how interventions will be coordinated across academic affairs, student affairs, and other institutional units (Jo et al., 2022). Policies should also address resource allocation, ensuring that support services have capacity to respond to the students identified as at-risk and that intervention intensity is calibrated to student need.

Fourth, institutions must establish policies governing evaluation and continuous improvement. Learning analytics is not a one-time implementation but an ongoing process that requires regular assessment and refinement. Policies should specify how the effectiveness of early warning systems and interventions will be evaluated, what metrics will be used to assess success, and how findings will inform system improvements. The goal should be continuous learning at the institutional level, mirroring the learning that analytics aims to support at the student level (Moreno-Marcos et al., 2025).

Fifth, institutions must develop policies governing faculty development and support. The successful implementation of learning analytics depends on faculty understanding, buy-in, and capacity to act on analytical insights. Policies should support professional development opportunities that help faculty interpret engagement data, design effective interventions, and integrate analytics into their teaching practice. Faculty should be partners in learning analytics, not merely subjects of it, and institutional policies should reflect this collaborative orientation (Alhothali et al., 2022).

The broader policy context for learning analytics extends beyond individual institutions to include system-level considerations. Governmental and accreditation bodies may develop standards for learning analytics practice, expectations for student data protection, and requirements for demonstrating that analytics investments improve outcomes. Professional organizations may develop ethical guidelines and best practice recommendations that inform institutional policy development. The findings of this study contribute to the evidence base that should inform these

broader policy discussions, demonstrating both the potential of learning analytics to support student success and the careful attention required to realize this potential responsibly.

In summary, the implications of this study span practical applications in early warning systems and teaching practice, as well as policy considerations for institutional adoption of learning analytics. Realizing the potential of learning analytics to improve student outcomes requires not only technical development of predictive models but also thoughtful attention to implementation, ethics, and the human dimensions of educational practice. The findings reported here provide empirical foundation for these efforts, supporting the development of data-informed approaches to online education that are both effective and respectful of the students they aim to serve.

8. Limitations

While this study provides valuable insights into the application of learning analytics for predicting student performance in online learning environments, several important limitations must be acknowledged. These limitations relate primarily to the scope of data collection, the range of variables examined, and the generalizability of findings across different contexts. Acknowledging these constraints is essential for interpreting the results appropriately and for identifying directions for future research that can address the gaps identified here (Alalawi et al., 2024).

Data Limited to One Institution

The most significant limitation of this study concerns the restriction of data collection to a single institution. All 347 student participants were drawn from one university, with data collected from courses delivered through a single Learning Management System implementation. This institutional specificity raises legitimate questions about the generalizability of findings to other educational contexts, including different types of institutions, different countries, and different online learning platforms (Jo et al., 2022).

Institutional context shapes online learning in numerous ways that may influence the relationships between engagement behaviors and academic performance. Institutional policies regarding course design, assessment practices, and student support services create particular conditions within which online learning occurs. The technological infrastructure available to students, including access to devices and reliable internet connectivity, varies across institutions and affects the feasibility and patterns of online engagement (Bashiru & Malgwi, 2026). The demographic composition of the student body, including factors such as age distribution, prior educational experience, and socioeconomic background, influences how students approach online learning and which engagement behaviors are most consequential for their success.

The single-institution design also limits sample size and diversity. While 347 students represents a reasonable sample for the analytical techniques employed, it constrains the statistical power available for detecting subgroup differences and for developing models that are robust across diverse student populations (Hubbard & Amponsah, 2025). Students in this sample were drawn from a limited range of academic disciplines and course types, potentially missing variation in engagement-performance relationships that might emerge across a broader range of educational contexts.

Furthermore, the study was conducted within a specific temporal context the post-pandemic period when online learning had become normalized but institutions were still adapting to sustained digital delivery. The engagement patterns observed and the predictive relationships identified may reflect this particular historical moment rather than stable features of online learning that would replicate across different time periods (Wu, 2026). As online learning continues to evolve, with new technologies, pedagogical approaches, and student expectations emerging, the predictive models developed here may require recalibration and validation in changed circumstances.

Limited Variables

A second important limitation concerns the range of variables included in the predictive models. While this study focused on behavioral indicators derived from Learning Management System logs login frequency, assignment submission rate and timing, forum participation, and learning resource usage these variables capture only a subset of the factors that influence student academic performance in online environments (Alhothali et al., 2022).

Notably absent from the models are variables related to student demographics, prior academic achievement, and psychological characteristics. Previous research has demonstrated that factors such as age, gender, socioeconomic status, prior grade point average, and self-reported motivation all contribute to predicting academic outcomes (Broadbent & Poon, 2023). The exclusion of these variables from the present study reflects the focus on behavioral indicators that can be collected unobtrusively from Learning Management Systems and that are potentially actionable for early intervention. However, this focus comes at the cost of model completeness, and the predictive accuracy achieved might have been higher if demographic and psychological variables had been incorporated.

The study also lacks information about the quality of student engagement, focusing instead on quantitative indicators of activity. Login frequency measures how often students access the platform but not what they do during those sessions or how attentively they engage with content. Forum participation counts posts but does not assess the cognitive depth or collaborative value of those contributions. Video viewing tracks completion rates but does not capture whether students are processing and retaining the content they view (Liu & Yu, 2023). These quantitative indicators serve as useful proxies for engagement, but they are imperfect measures that may miss important qualitative dimensions of the learning experience.

Course-level variables that might moderate the relationships between engagement and performance were also not systematically examined. Course characteristics such as disciplinary field, level of difficulty, instructional quality, assessment design, and the extent to which engagement is required or optional likely influence which behaviors matter most for success (Moreno-Marcos et al., 2025). The present study treated course as a context for data collection rather than as a variable to be analyzed, potentially obscuring important variation in how engagement-performance relationships operate across different course types.

The temporal granularity of analysis represents another limitation. While the study examined engagement patterns across weeks and identified early predictive signals, the analysis did not fully exploit the rich temporal structure of the log data. More sophisticated time-series approaches might reveal finer-grained patterns of engagement such as sequences of activities within sessions, transitions between different types of learning tasks, or responsiveness to specific course events that could enhance predictive accuracy and provide deeper insight into learning processes (Qiu et al., 2022).

Finally, the study's focus on final course grades as the sole outcome measure, while practically justified, provides an incomplete picture of student success. Grades reflect institutional assessment of learning but do not capture other important outcomes such as skill development, knowledge retention, satisfaction, persistence in subsequent courses, or degree completion. Different outcome measures might show different patterns of relationship with engagement behaviors, and the predictive models developed here may not generalize to these alternative conceptualizations of success (Wong et al., 2021).

These limitations do not invalidate the study's findings but rather contextualize them, indicating the boundaries within which conclusions can be confidently drawn and highlighting opportunities for future research to extend and refine understanding of learning analytics for predicting student performance in online environments.

9. Future Research Directions

While this study contributes meaningful insights to the growing body of learning analytics research, it also illuminates several pathways for future investigation that can address the limitations identified and extend understanding of how predictive analytics can support student success in online learning environments. These future research directions encompass methodological advancements, theoretical development, and practical applications that together can advance the field toward more robust, generalizable, and impactful implementations (Alalawi et al., 2024).

Cross-Institution Datasets

The limitation of data collection to a single institution represents a significant constraint on the generalizability of findings and suggests an urgent need for research employing cross-institution datasets. Future studies should seek to aggregate data from multiple institutions representing diverse contexts different countries, institutional types, student populations, and Learning Management System platforms to examine whether the predictive relationships identified here replicate across settings and to identify contextual factors that moderate these relationships (Jo et al., 2022).

Cross-institution research would enable examination of how institutional characteristics shape the engagement-performance relationship. Institutions vary in their student demographics, admission standards, academic support resources, and technological infrastructure, all of which may influence how students engage with online learning and which behaviors most strongly predict success (Bashiru & Malgwi, 2026). Understanding these contextual variations is essential for developing predictive models that can be confidently deployed across diverse institutional settings and for identifying which findings reflect universal features of online learning and which are context-dependent.

Such research would also enable larger and more diverse samples, supporting more sophisticated analytical approaches and more robust model development. With data from multiple institutions, researchers could employ techniques such as hierarchical modeling to partition variance in outcomes into student-level and institution-level components, examining how much of the predictive power of engagement behaviors is consistent across contexts versus specific to particular institutional settings (Hubbard & Amponsah, 2025). Larger samples would also support more fine-grained analysis of subgroup differences, including examination of whether predictive relationships vary by student demographic characteristics, academic preparation, or disciplinary focus.

Cross-institution datasets would facilitate the development of more generalizable early warning systems that could be implemented across institutions with confidence in their performance. Currently, institutions considering learning analytics adoption must either develop their own models from scratch or rely on findings from potentially non-comparable contexts. Collaborative research networks that pool data across institutions could develop and validate models that are robust across diverse settings, accelerating the translation of research into practice and supporting more widespread adoption of evidence-based early warning systems (Moreno-Marcos et al., 2025).

The practical challenges of cross-institution research are substantial, including issues of data harmonization, privacy protection, and inter-institutional collaboration. Different Learning Management Systems capture and store data in different formats, requiring careful work to map variables across platforms and ensure comparability. Privacy regulations and institutional policies governing student data vary across jurisdictions, necessitating careful attention to data sharing agreements and ethical protocols. Despite these challenges, the potential benefits of cross-institution research justify sustained effort to overcome these obstacles and build the collaborative infrastructure necessary for such investigations (Wu, 2026).

AI-Based Adaptive Learning Systems

A second major direction for future research concerns the integration of predictive analytics with AI-based adaptive learning systems that can respond dynamically to student engagement patterns.

The present study demonstrates that student behaviors predict performance, but it does not address how these predictions might be used to trigger adaptive responses within the learning environment itself. Future research should explore the development and evaluation of systems that use real-time behavioral data to personalize learning experiences, providing tailored content, recommendations, and support based on individual student engagement patterns (Liu & Yu, 2023).

Adaptive learning systems represent the logical extension of predictive analytics from identification to intervention. Rather than simply flagging at-risk students for human intervention, these systems would automatically adjust the learning environment in response to detected engagement patterns. For example, a student showing declining login frequency might receive automated messages suggesting study strategies or offering connections to support resources. A student struggling with video content might be presented with alternative explanations or additional practice opportunities. A student who consistently submits work at the last minute might receive prompts encouraging earlier planning and breaking assignments into smaller steps (Alhothali et al., 2022).

The development of such systems requires research addressing multiple interconnected questions. What types of adaptive responses are most effective for different engagement patterns? How should systems balance automation with human judgment, ensuring that adaptive responses support rather than supplant instructor-student relationships? How can systems be designed to respect student autonomy and avoid creating experiences that feel manipulative or overly deterministic? These questions require interdisciplinary collaboration bringing together learning analytics researchers, instructional designers, human-computer interaction specialists, and educational psychologists (Broadbent & Poon, 2023).

Research on adaptive systems must also address questions of timing and intensity. How early in a course should adaptive responses be triggered? How intensive should interventions be, and how should intensity escalate if initial responses prove insufficient? What combinations of adaptive responses are most effective for different student profiles? Answering these questions requires longitudinal research designs that track student responses to adaptive interventions over time, examining both immediate behavioral changes and longer-term impacts on learning outcomes (Wong et al., 2021).

The ethical dimensions of AI-based adaptive learning systems warrant careful research attention. As systems become more sophisticated in their ability to detect engagement patterns and adapt learning experiences, questions arise about transparency, consent, and the potential for unintended consequences. Students should understand how their data is being used to shape their learning experiences and should have meaningful opportunities to opt out or influence system behavior. Research should examine student perceptions of adaptive systems, identifying design principles that promote trust, engagement, and perceived autonomy (Qiu et al., 2022).

Finally, research should examine the integration of adaptive systems with human support structures, exploring models that combine automated responses with human intervention in complementary ways. The most effective approaches may be those that use AI to handle routine monitoring and initial responses, freeing human instructors and advisors to focus on students with the most complex needs and to provide the uniquely human elements of teaching empathy, inspiration, and genuine relationship that technology cannot replicate (Anthonysamy et al., 2020).

In summary, future research should extend the findings of this study in two complementary directions: outward to encompass cross-institution datasets that test generalizability and identify contextual moderators, and inward to develop AI-based adaptive systems that translate predictive insights into personalized learning experiences. Together, these directions promise to advance both scientific understanding of online learning processes and practical capacity to support student success through data-informed, personalized education.

10. Conclusion

This study set out to investigate the potential of learning analytics for predicting student academic performance in online learning environments, with the dual aims of advancing theoretical understanding of engagement-performance relationships and providing practical guidance for early intervention and instructional improvement. Across the preceding chapters, we have examined the landscape of online education, reviewed the relevant literature, developed a conceptual framework grounded in self-regulated learning theory, employed rigorous quantitative methods to analyze student interaction data, and interpreted the findings in light of their implications for practice and policy. This conclusion synthesizes the key insights emerging from this investigation, reflecting on the importance of learning analytics, summarizing the key predictors of student performance identified, and articulating the potential benefits for online education that justify continued investment in this approach (Alalawi et al., 2024).

The Importance of Learning Analytics

The rapid expansion of online learning documented at the outset of this study has fundamentally transformed the educational landscape, creating both unprecedented opportunities for access and flexibility and significant challenges for student support and success. As institutions have moved substantial portions of their educational offerings online, they have simultaneously generated vast quantities of data about how students engage with learning data that, until recently, remained largely unexamined and unutilized (Liu & Yu, 2023). Learning analytics emerges from this context as both a necessary response to the challenges of online education and a natural outgrowth of the digital infrastructure that now supports teaching and learning.

The importance of learning analytics lies in its capacity to transform raw data into actionable intelligence. The Learning Management Systems that have become ubiquitous in higher education Moodle, Canvas, Blackboard, and others function not only as platforms for content delivery and assessment but as sophisticated data collection instruments, capturing detailed records of student activity that can be analyzed to reveal patterns, predict outcomes, and guide intervention (Hubbard & Amponsah, 2025). This study has demonstrated that these data, when appropriately processed and analyzed, contain meaningful signals about student progress and potential difficulty, signals that can be detected early enough to enable timely support.

The theoretical grounding of this study in self-regulated learning theory has provided a framework for interpreting behavioral data in terms of underlying psychological processes. Login frequency, assignment submission timing, forum participation, and learning resource usage are not merely arbitrary indicators but manifestations of the planning, monitoring, and strategic engagement that characterize effective self-regulated learners (Broadbent & Poon, 2023). This theoretical connection elevates learning analytics beyond purely empirical prediction, connecting observed behaviors to established understanding of how learning occurs and enabling theoretically informed intervention design.

The importance of learning analytics is particularly acute in the context of developing countries, where resource constraints, large class sizes, and limited academic support systems compound the challenges of online education. The present study, situated in such a context, contributes to addressing the research gap identified in the literature review regarding the limited attention to learning analytics in developing country settings (Bashiru & Malgwi, 2026). The findings demonstrate that predictive models developed in these contexts can achieve accuracy comparable to those reported in studies from wealthier nations, suggesting that learning analytics is not merely a luxury for well-resourced institutions but a potentially valuable tool for improving educational equity and outcomes across diverse settings.

Key Predictors of Student Performance

The empirical findings of this study have identified a set of key behavioral predictors that reliably distinguish successful from struggling students in online learning environments. These predictors, ranked by their importance in the random forest model, provide a focused set of indicators that educators and institutions can monitor to identify students in need of support and to understand the engagement patterns that contribute to academic success.

Assignment submission rate emerged as the single most powerful predictor of student performance, confirming that consistent task completion is fundamental to success in online learning. Students who submit a high proportion of assigned work demonstrate the goal attainment and task persistence that characterize effective self-regulation, and their consistent engagement provides multiple opportunities for feedback and learning that accumulate across the course (Qiu et al., 2022). The practical implication is straightforward: monitoring submission rates provides an early warning system that identifies students at risk before they miss so many assignments that recovery becomes impossible.

Login frequency and consistency ranked second in importance, revealing that sustained, regular engagement with the learning environment matters as much as or more than total time spent. Students who access their courses consistently across weeks establish learning routines that support distributed practice, maintain connection with course developments, and signal ongoing motivational investment (Jo et al., 2022). The finding that struggling students often exhibit declining login frequency before they miss assignments suggests that monitoring platform access can provide even earlier warning than tracking submission rates alone.

Video lecture completion rate ranked third, highlighting the central role of content engagement in online learning success. Students who watch videos thoroughly, persisting through complete lectures rather than sampling selectively, achieve better outcomes, suggesting that the depth and quality of content engagement matters alongside its quantity. This finding underscores the importance of designing video content that sustains attention and of supporting students in developing effective strategies for learning from video (Liu & Yu, 2023).

Assignment submission timeliness ranked fourth, providing predictive information beyond submission rate alone. Students who submit work early or consistently meet deadlines demonstrate planning and time management capabilities that contribute to success, while last-minute submission patterns signal potential difficulties even when assignments are completed. This finding suggests that interventions might usefully target not only whether students complete work but when they complete it, helping students develop the planning skills that support sustained engagement (Anthonysamy et al., 2020).

Forum participation ranked fifth, confirming that social engagement contributes to success but may be less central than engagement with assessments and content. The variation in importance across course contexts suggests that the role of forum participation depends on course design, with social engagement mattering more in courses where collaborative learning is central to the pedagogical approach (Wong et al., 2021).

Potential Benefits for Online Education

The potential benefits of learning analytics for online education, as illuminated by this study, operate at multiple levels supporting individual students, empowering instructors, informing institutional strategy, and ultimately advancing the quality and equity of digital learning.

For students, the most direct benefit lies in the possibility of earlier and more effective support when difficulties arise. The traditional model of academic intervention, which waits until poor performance on high-stakes assessments reveals problems, often intervenes too late to alter trajectories. Early warning systems based on behavioral indicators can identify at-risk students weeks or even months before final assessments, creating windows of opportunity for support that can make the difference between failure and success (Moreno-Marcos et al., 2025). For students who would

otherwise struggle silently, disengaging gradually until they reach a point of no return, learning analytics offers the possibility of being seen, being reached, and being supported before it is too late.

For instructors, learning analytics provides visibility into student engagement that would otherwise be unavailable. In traditional classrooms, instructors can read nonverbal cues, notice who is paying attention, and sense the collective energy of the room. In online environments, these sources of information disappear, leaving instructors teaching into what often feels like a void. Engagement dashboards that summarize login patterns, submission rates, and forum participation give instructors windows into student activity, enabling them to identify students who may need outreach and to assess how the class as a whole is engaging with course materials (Alhothali et al., 2022). This visibility supports more responsive teaching and more meaningful connection with students.

For institutions, learning analytics offers the potential for more strategic allocation of support resources. Rather than distributing support evenly across all students or relying on self-referral that may miss those most in need, institutions can target intervention resources toward students identified as at-risk by predictive models. This targeting increases the efficiency and effectiveness of support services, concentrating effort where it is most needed and most likely to make a difference (Wu, 2026). Over time, aggregate analytics can identify patterns of engagement and difficulty across courses and programs, informing curriculum review, instructional development, and resource allocation at the institutional level.

For the field of online education more broadly, learning analytics contributes to the ongoing professionalization and evidence-based improvement of digital teaching and learning. By making visible the relationships between engagement behaviors and outcomes, analytics provides empirical grounding for pedagogical decisions that might otherwise be guided by intuition, tradition, or ideology. Course designs can be evaluated based on their effects on engagement and learning. Teaching practices can be refined based on evidence about what works. The accumulation of findings across studies, including this one, builds a knowledge base that can guide the continued development of online education as a mature and effective form of teaching and learning (Alalawi et al., 2024).

In conclusion, this study has demonstrated that learning analytics, grounded in theoretical understanding of self-regulated learning and implemented through rigorous quantitative methods, can generate accurate predictions of student performance in online learning environments. The key predictors identified assignment submission rate, login frequency, video completion, submission timeliness, and forum participation provide a focused set of indicators that can inform early warning systems, instructional practice, and course design. The potential benefits for students, instructors, institutions, and the field of online education justify continued investment in learning analytics research and development, always with careful attention to the ethical considerations and human dimensions that must accompany any application of data to education. As online learning continues to expand and evolve, the insights generated through learning analytics will become increasingly essential for fulfilling the promise of digital education: expanded access, improved outcomes, and more equitable opportunities for all learners.

References

- Alalawi, K., Athauda, R., & Chiong, R. (2024). An extended learning analytics framework integrating machine learning and pedagogical approaches for student performance prediction and intervention. *International Journal of Artificial Intelligence in Education*, 34. <https://doi.org/10.1007/s40593-024-00429-7>
- Alhothali, A., Albsisi, M., Assalahi, H., & Aldosemani, T. (2022). Predicting student outcomes in online courses using machine learning techniques: A review. *Sustainability*, 14(10), 6199. <https://doi.org/10.3390/su14106199>
- Anthonyamy, L., Koo, A. C., & Hew, S. H. (2020). Self-regulated learning strategies in higher education: Fostering digital literacy for sustainable lifelong learning. *Education and Information Technologies*, 25(4), 2393-2414. <https://doi.org/10.1007/s10639-020-10201-8>

- Bashiru, S., & Malgwi, Y. M. (2026). A machine learning-based early warning system for identifying at-risk university students in Nigerian higher education. *MJSE*, 1-15. <http://oer.tsuniversity.edu.ng/index.php/mjse/article/view/1759>
- Broadbent, J., & Poon, W. L. (2023). Self-regulated learning strategies in online learning environments: A systematic review. *The Internet and Higher Education*, 57, 100890. <https://doi.org/10.1016/j.iheduc.2022.100890>
- Clarín, A. S., & Baluyos, E. L. (2022). Challenges encountered in the implementation of online distance learning. *EduLine: Journal of Education and Learning Innovation*, 2(1), 33–46. <https://doi.org/10.35877/454RI.eduline591>
- Hubbard, K., & Amponsah, S. (2025). Feature engineering on LMS data to optimize student performance prediction. *arXiv*. <https://arxiv.org/abs/2504.02916>
- Jo, I. H., Park, Y., & Lee, H. (2022). Learning analytics in higher education: A review of the literature from 2012 to 2022. *Educational Technology Research and Development*, 70(3), 987-1015. <https://doi.org/10.1007/s11423-022-10106-4>
- Liu, M., & Yu, D. (2023). Towards intelligent E-learning systems. *Education and Information Technologies*, 28(7), 7845–7876. <https://doi.org/10.1007/s10639-022-11479-6>
- Moreno-Marcos, P. M., et al. (2025). Integration of multiple sources to anticipate student performance using learning analytics. In *2025 13th International Conference on Education Technology and Computers*. IEEE. <https://ieeexplore.ieee.org/document/11194780>
- Qiu, F., et al. (2022). Predicting students' performance in e-learning using learning process and behaviour data. *Scientific Reports*, 12(1), 453. <https://doi.org/10.1038/s41598-021-03867-8>
- Stratview Research. (2026). *E-learning market trend, share & forecast 2022-2026*. <https://www.stratviewresearch.com/2748/e-learning-market.html>
- Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2021). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human-Computer Interaction*, 37(4), 309-327. <https://doi.org/10.1080/10447318.2020.1825144>
- Wu, Y. (2026). Academic performance in the digital age: Rethinking student success through digital learning ecosystems. In J. R. Hermann (Ed.), *Academic performance - Student success in the transformative digital age*. IntechOpen. <https://doi.org/10.5772/intechopen.1014529>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.