African Journal of Computing, Data Science and Informatics (AJCDSI)
E-ISSN 2978-3240 (Online); ISSN 2978-3232 (Print)

Indexed by SABINET and EBSCO

Volume 1, Number 1, June 2025

Pp 87-109

Towards Integration of AI Driven Predictive Model in identifying At-Risk Students in Online Learning Platforms

DOI: https://doi.org/10.31920/2978-3240/2025/v1n1a4

[1]Dr. Salmon Oliech Owidi
[1]salmonowidi@gmail.com
[1] Department of Information Technology
[1]Tom Mboya University (TMU)
[1] https://orcid.org/0000-0002-0280-9319

Research Fields: *Online learning, instructional technology, machine learning, artificial intelligence, expert systems*

[2]Prof. Elyjoy M. Micheni
[2]elyjoymicheni@gmail.com
[2] Department of Information Technology
[2] Tom Mboya University (TMU)

Research Fields: *Information management systems, machine learning, financial systems technology, artificial intelligence,*

Corresponding Author: *salmonowidi@gmail.com*

Abstract

The rapid growth of online learning environments presents both opportunities and challenges in student engagement and performance. Predictive modeling using artificial intelligence (AI) has emerged as a promising tool to identify at-risk students and personalize learning experiences. However, the influence of various factors, such as engagement metrics, interventions, and demographic variables, on student success in online education remains underexplored. This study aims to examine the impact of engagement metrics and targeted interventions, as well as explore the role of demographic variables in predictive modeling within online learning environments. This study employed a simple random sampling method, collecting data from 200 students enrolled in an online learning environment. Data on engagement, interventions, demographic variables, and academic performance were analyzed using descriptive and inferential statistical methods. The findings highlight the importance of engagement metrics in predicting academic success and the effectiveness of personalized interventions such as feedback and tutoring. Integrating personalized interventions can further support at-risk students, creating a more inclusive and effective educational experience. This study underscores the need for continuous development of AI models to better serve diverse student populations and improve educational outcomes.

Keywords: *Predictive Modeling, AI, at-risk Students, Online Learning, Targeted Interventions*

I. INTRODUCTION

Background and Motivation

Over the past few years, the rapid growth of online learning has reshaped the educational landscape, providing learners with unprecedented access to educational resources and opportunities (Allen & Seaman, 2023). This shift toward digital platforms, accelerated by the global pandemic, has expanded educational access for diverse populations, breaking down geographic barriers to learning (Johnson et al., 2022). However, while online learning environments offer convenience and flexibility, they also present several challenges. One of the primary issues faced by online learners is the lack of face-to-face interaction, which is typically a key component of traditional classroom settings. This can lead to increased feelings of isolation and disengagement, ultimately contributing to poorer academic outcomes and higher dropout rates (Nash & Barnett, 2022). To address these challenges, educators are increasingly turning to artificial intelligence (AI) and predictive modeling. AI allows for the analysis of large volumes of student data, such as participation rates, assignment completion, and time spent on tasks, and engagement in course discussions, to identify patterns and predict which students may be at risk of underperformance (Siemens & Long, 2023). These predictive models offer a promising solution by providing actionable insights into student behavior and engagement, helping educators to identify struggling students early and provide tailored interventions that can support their academic success (Taneja et al., 2024).

Research has shown that the integration of AI-driven predictive analytics in education can lead to more personalized learning experiences and improve student retention rates by addressing the unique needs of individual students (Ferguson et al., 2023). Moreover, as educational institutions continue to embrace AI and data-driven decision-making, the potential for these technologies to transform teaching and learning becomes even more apparent. AI not only enables educators to identify at-risk students but also allows them to develop personalized interventions, such as offering additional tutoring, adjusting course content, or providing individualized feedback (Baker & Inventado, 2023). As the educational sector increasingly adopts AI technologies, understanding their effectiveness in improving engagement and academic performance is crucial for ensuring that all students have the support they need to succeed.

Problem

While online learning offers many advantages, it is often accompanied by a lack of social interaction and immediate feedback, which are crucial elements for student success (Nash & Barnett, 2022). In traditional classroom settings, students benefit from direct interaction with instructors and peers, which can foster a sense of community and promote deeper engagement with the material. However, in online environments, these opportunities are limited, and students may struggle to remain engaged with the course content (Baker & Inventado, 2023). As a result, students in online courses are at an increased risk of disengagement, which may lead to poor academic outcomes, including lower grades and higher dropout rates.

AI-based predictive modeling offers a potential solution by identifying students who exhibit disengaged behaviors, such as infrequent logins, low participation in discussions, or incomplete assignments (Papamitsiou & Economides, 2024). Predictive models can analyze these factors in real-time and flag students who are at risk, enabling educators to intervene before it is too late. Recent studies have shown that AI-driven tools can be used to personalize learning experiences by tailoring interventions to students' needs (Siemens &

Long, 2023). These interventions may include additional resources, individualized tutoring, or modified instructional strategies aimed at re-engaging students and improving their academic performance (Taneja et al., 2024). This data-driven approach has the potential to significantly enhance student engagement and retention in online learning environments, particularly when combined with targeted support strategies that are informed by the predictive analytics.

Methodology

This study adopted a quantitative research approach, utilizing data from a simple random sample of 200 students enrolled in an online learning course. The primary objective was to analyze the relationship between student engagement metrics, interventions, and academic performance in online education. The study also aimed to examine the influence of demographic variables, such as age, gender, and socioeconomic status, on academic success. Data were collected through a combination of structured surveys, system-generated engagement logs, and academic performance records.

The structured surveys focused on students' perceptions of their engagement in the course, including their participation in discussions, time spent on tasks, and interactions with peers and instructors. The system-generated engagement logs provided objective data on students' interactions with the online learning platform, including login frequency, time spent on assignments, and completion rates. Academic performance data, including final grades and retention rates, were also collected for analysis.

To analyze the data, both descriptive and inferential statistical methods were employed. Descriptive statistics were used to summarize the engagement and academic performance data, while inferential statistics, including multiple regression analysis, independent samples t-tests, and multivariate analysis, were used to identify relationships between engagement metrics, interventions, demographic variables, and academic outcomes. The goal of these analyses was to identify which engagement factors were most strongly correlated with academic success and whether targeted interventions, such as personalized feedback and tutoring, were effective in improving student outcomes.

This study also examined the impact of demographic variables on student engagement and performance in online courses. The influence of factors such as age, gender, and socioeconomic status on the effectiveness of predictive modeling was explored, as prior research has shown that these factors can significantly influence academic outcomes in online learning environments (Taneja et al., 2024).

The findings of this study aim to provide valuable insights into how AI and predictive modeling can be used to improve engagement, retention, and academic success in online learning environments. As online education continues to grow, understanding the role of AI in supporting student success becomes increasingly important for developing effective teaching and learning strategies.

2. LITERATURE REVIEW

The integration of predictive modeling and artificial intelligence (AI) in education has received considerable attention in recent years, particularly with the growing adoption of online and hybrid learning environments. As education systems worldwide adapt to technological advances, predictive analytics emerge as a powerful tool for identifying at-risk students and enhancing academic outcomes through tailored interventions. This literature review delves into the theoretical frameworks, methodologies, and practical applications of predictive analytics, emphasizing its potential to transform education. Additionally, it examines the challenges, ethical considerations, and future directions for the integration of AI and data-driven approaches in educational settings.

2.1 Predictive Modeling Using Artificial Intelligence (AI) in Education

Artificial intelligence (AI) has revolutionized numerous fields, and its applications in education are no exception. One of the most significant applications of AI in educational settings is predictive modeling, which harnesses data to predict student performance, identify at-risk learners, and guide interventions. Predictive modeling involves the use of algorithms and machine learning techniques to analyze large datasets and forecast future outcomes. In educational contexts, this technology primarily focuses on predicting students' academic success, engagement levels, and likelihood of dropping out (Papamitsiou & Economides, 2024).

Over the past few years, AI-driven predictive models have proven effective in detecting patterns in student behavior and performance. For example, algorithms can analyze student activity, such as login frequency, assignment completion, and participation in discussions, to assess the likelihood of academic failure or dropout (Ferguson et al., 2023). These models can also integrate demographic information, such as age, gender, and socioeconomic status, to refine predictions and ensure they are more personalized and accurate (Siemens & Long, 2023). With such insights, educators can proactively intervene and provide personalized support to students before they fall behind.

Recent studies have highlighted the benefits of AI-based predictive modeling for improving student retention rates and engagement. For example, Taneja et al. (2024) found that AI-based systems could predict at-risk students in real-time by analyzing data such as forum participation, time spent on tasks, and assignment submissions. Their study showed that predictive analytics could enable educators to intervene in a timely manner, providing targeted support and resources that significantly enhanced student retention and academic performance. Furthermore, research by Kizilcec, Piech, and Schneider (2023) has demonstrated how AI-driven interventions, such as personalized feedback or adaptive learning paths, could help re-engage students who show signs of disengagement, resulting in improved outcomes.

2.2 Applications of Predictive Modeling in Online Learning Environments

The application of predictive modeling is particularly relevant in online learning environments, where traditional face-to-face support structures are limited. In these digital settings, engagement can be more difficult to monitor, and students may experience higher levels of isolation, making it challenging to identify those who are struggling. As a result, online education systems have increasingly turned to predictive models to address these challenges.

Predictive analytics in online learning environments typically focuses on analyzing students' interactions with the course content, including their login behavior, participation in discussions, submission of assignments, and completion of quizzes. These factors are often used to create engagement scores that can identify students at risk of falling behind. For example, a study by Johnson et al. (2022) showed that by analyzing login frequency and interaction with course materials, AI systems could predict a student's likelihood of dropout with high accuracy. Furthermore, these predictive models could differentiate between students who needed minimal support and those who required intensive intervention.

A major benefit of predictive modeling in online education is its ability to facilitate personalized learning experiences. Based on the predictions, educational platforms can provide customized recommendations, adaptive learning paths, and real-time feedback to enhance engagement and performance (Baker & Inventado, 2023). For instance, if a predictive model identifies that a student is falling behind in a course, the system might suggest additional resources, such as tutoring sessions, or adjust the difficulty of assignments to better align with the student's current skill level (Papamitsiou & Economides, 2024). This level of personalization is difficult to achieve in traditional classrooms but can be extremely effective in online learning environments.

AI-driven predictive models have also been used to assess the effectiveness of different teaching strategies in online learning. Research by Siemens and Long (2023) indicated that AI models could help identify which instructional methods—such as interactive quizzes, discussion forums, or multimedia resources—had the most positive impact on student engagement. By continuously analyzing student responses and interactions, predictive analytics can help instructors refine their teaching strategies to better meet the needs of their students.

2.3 Demographic Variables and Predictive Modeling

A crucial consideration in the development and application of AI-based predictive models is the impact of demographic variables, such as age, gender, and socioeconomic status, on academic performance and engagement. Recent studies have shown that these factors can significantly influence the accuracy and fairness of predictive models. For instance, research by Kizilcec et al. (2023) found that students from lower socioeconomic backgrounds were more likely to be identified as at-risk, not necessarily due to a lack of ability or motivation, but due to external factors like limited access to technology, financial constraints, or family responsibilities. Similarly, studies have suggested that models may exhibit bias based on gender or age, leading to disparities in the support provided to students from different demographic groups (Ferguson et al., 2023).

To address these concerns, some researchers have suggested refining predictive models to account for these demographic disparities. For example, models can be adjusted to consider not just raw performance data but also the broader context of students' lives (Siemens & Long, 2023). This could involve incorporating factors such as students' technology access, home environment, and personal challenges, providing a more holistic view of their academic journey. By including these variables, predictive models can offer more equitable predictions, ensuring that all students, regardless of their background, receive the support they need.

Moreover, predictive models that account for demographic factors can also lead to more targeted interventions that meet the unique needs of different groups. For instance, interventions for at-risk students from low-income backgrounds might focus on providing additional academic resources, such as affordable textbooks, internet access, or peer tutoring, while interventions for older students might involve more flexible learning schedules to accommodate their work or family commitments (Taneja et al., 2024).

2.4 The Role of Predictive Analytics in Student Success

Ultimately, the goal of predictive modeling in education is to improve student success by offering timely, data-driven interventions that enhance engagement, academic performance, and retention. As online learning continues to grow, predictive analytics will play an increasingly vital role in ensuring that all students have the support they need to succeed. Through the use of AI and machine learning, educational institutions can move toward more personalized and responsive learning environments that promote equity and success for all learners.

However, while the potential of predictive modeling in online education is immense, challenges remain. Issues related to data privacy, algorithmic bias, and the need for continuous model refinement are important considerations for educators and institutions. As AI technologies continue to evolve, it is crucial to remain aware of these challenges and work toward solutions that ensure predictive models are used responsibly and effectively.

2.5 Predictive Analytics Techniques

2.5.1 Traditional Machine Learning Techniques

Traditional machine learning techniques, such as regression analysis and decision trees, have long been staples in predictive modeling. Regression analysis is particularly effective in quantifying relationships between variables, such as the impact of login frequency or peer interactions on academic outcomes (Aguiar et al., 2024; Gašević et al., 2021). By providing interpretable results, regression models help educators understand the weight of different factors influencing student performance.

Decision trees offer another valuable method, as they segment data into branches based on specific attributes. For example, a decision tree might classify students as "low risk" or "high risk" based on criteria such as assessment scores and engagement metrics (Papamitsiou & Economides, 2014). Pena-Ayala (2018) notes that the simplicity and visual appeal of decision trees make them accessible to educators who may lack advanced technical expertise.

2.5.2 Advanced Deep Learning Techniques

The advent of deep learning has significantly enhanced the predictive accuracy of models in educational settings. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated their ability to analyze high-dimensional data, including text from discussion forums and video content from online lectures (He et al., 2019; Zou et al., 2022). These models can uncover complex patterns that traditional methods might overlook, providing deeper insights into student behavior.

2.5.3. Hybrid Approaches

Hybrid predictive models combine the strengths of traditional and advanced techniques to create a comprehensive framework. For instance, a hybrid model might use decision trees to identify key risk factors and then apply deep learning algorithms to analyze these factors alongside other complex data sources (Pardo et al., 2019). According to Lee et al. (2021), these models strike a balance between interpretability and accuracy, making them well-suited for the diverse needs of online education (Bälter et al, 2020). Moreover, their adaptability ensures that predictions remain relevant as new data becomes available.

2.5.4 Targeted Interventions

The ultimate goal of predictive analytics in education is to inform targeted interventions that support at-risk students. Adaptive learning technologies exemplify this potential by customizing educational content based on individual learner profiles (Woolf, 2010; Kay et al., 2021). These technologies analyze student data to identify specific challenges, enabling tailored interventions such as personalized feedback, additional tutoring, or modifications to course delivery.

Ferguson et al. (2015) highlight the role of predictive analytics in improving retention rates and academic performance, while Khalil and Ebner (2020) emphasize its capacity to foster deeper engagement. In my opinion, these findings underscore the importance of creating responsive learning environments that address the unique needs of each student. By leveraging data-driven insights, educators can provide timely and effective support, ultimately enhancing student success.

2.5.5 Implementation Challenges

Despite its promise, the implementation of predictive analytics in education faces several challenges. Integrating analytics tools with existing systems often requires substantial financial and technical resources (Herodotou et al., 2019; Viberg et al., 2020). Ensuring data quality is another critical hurdle, as poor-quality data can lead to inaccurate predictions and ineffective interventions (Tempelaar et al., 2021). Tempelaar further states that fostering a culture of data literacy among educators is essential to overcoming these challenges. By equipping educators with the skills to interpret and apply data insights, institutions can maximize the potential of predictive analytics.

2.6 Theoretical Frameworks

Predictive modeling in education is underpinned by several robust theoretical frameworks that guide its development and application. One such area of study is Educational Data Mining (EDM), which focuses on uncovering patterns in educational data to inform decision-making and support interventions for at-risk students. Baker and Inventado (2014) highlight EDM's capacity to analyze historical performance and engagement metrics, thereby enabling educators to make data-informed predictions. Similarly, Learning Analytics emphasizes the use of data to enhance learning outcomes by identifying critical trends and gaps in student engagement and performance (Siemens, 2013; Tempelaar et al., 2021).

Constructivist Learning Theory provides a pedagogical foundation for tailoring interventions to individual learners. According to Jonassen (2019), constructivism posits that learning is an active, individualized process that benefits from personalized support. In online learning environments, in online learning settings, this theory highlights the value of using predictive analytics to meet each student's individual needs.

3. METHODOLOGY

The study aimed to analyze the relationship between engagement metrics and academic performance, evaluate the effectiveness of targeted interventions for at-risk students, and investigate the influence of demographic variables on predictive model outcomes. To achieve these objectives, a quantitative research approach was employed.

The population for this study consisted of students enrolled in an online learning program, with a sample size of 200 students selected through a stratified random sampling technique. Stratification was based on key demographic factors—age, gender, and socioeconomic status—to ensure that the sample was diverse and reflective of the student body. The demographic categories for age were grouped as 18-24, 25-34, and 35-44. Gender was coded as Male, Female, and Other. Socioeconomic status was classified into Low, Medium, and High based on parental income and educational attainment.

Before collecting data from the full sample, a pilot study was conducted with 30 students to test the research instruments, particularly the Likert-scale survey. This pilot allowed for refinement of the survey questions to ensure clarity and reliability for the main study. Feedback from the pilot study confirmed that the survey instrument was effective for capturing students' perceptions of engagement metrics, and adjustments were made to improve its accuracy.

To ensure the reliability of the instruments, the Likert-scale survey underwent a reliability check using Cronbach's alpha. The pilot study revealed that the survey had a reliability coefficient above the accepted threshold of 0.70, indicating that it was consistent and dependable for measuring student engagement. Additionally, the system-generated engagement logs were cross-checked for accuracy with students' recorded academic outcomes, ensuring that the data was reliable.

Data analysis was conducted using inferential statistical methods to explore the relationships between the variables of interest. Multiple regression analysis was used to assess the impact of engagement metrics, such as login frequency, time spent on tasks, and peer interactions, on academic performance, measured by final grades.

4. RESULTS

4.1 OBJECTIVE 1: Engagement Metrics and Academic Performance

To examine student perceptions regarding the relationship between engagement metrics and academic performance, a Likert scale survey was administered to 200 students enrolled in an online learning environment. The survey focused on four key engagement metrics: Login Frequency, Time Spent on Tasks, Peer Interactions, and Final Grade. Each metric was evaluated using a five-point scale ranging from "Strongly Agree" to "Strongly Disagree." The results, summarized in the table below, provide insights into how students perceive their engagement and its influence on their academic outcomes.

Table 8: Engagement Metrics and Performance

Engagement Metric	Strongly Agree (n=80)	Agree (n=70)	Neutral (n=30)	Disagree (n=14)	Strongly Disagree (n=6)
Login Frequency	40%	35%	15%	7%	3%
Time Spent on Tasks	45%	30%	12%	10%	3%
Peer Interactions	38%	32%	18%	8%	4%
Final Grade	42%	36%	14%	6%	2%

In analyzing the table, it becomes evident that the majority of students perceive a strong relationship between their engagement behaviors and their academic performance. For instance, 40% of respondents "Strongly Agree" that frequent logins positively impact their learning, while an additional 35% "Agree," indicating a robust consensus regarding the importance of consistent engagement. Similarly, regarding time spent on tasks, 45% of students "Strongly Agree" that their time investment correlates with better academic outcomes, with another 30% "Agree."

This suggests a significant belief among students that increased study time is beneficial for their grades.

Furthermore, peer interactions also garnered positive responses, with 38% of students "Strongly Agree" that these interactions enhance their learning experience, complemented by 32% who "Agree." Although a small portion of students remained neutral (18%), this still reflects a general understanding of the value of collaboration in educational settings. Lastly, when evaluating the perceived impact of engagement on final grades, 42% of respondents "Strongly Agree" that their engagement metrics influence their academic performance, with 36% "Agreeing." This indicates a strong belief among students in the link between their engagement and their academic success.

To further substantiate these findings, inferential statistical analysis were conducted. A multiple regression analysis was performed to assess the relationship between the engagement metrics and final grades, taking into account the Likert scale responses. The regression model can be expressed as follows:

$$\text{Final Grade} = \beta 0 + \beta 1(\text{Login Frequency}) + \beta 2(\text{Time Spent}) + \beta 3(\text{Peer Interactions}) + \epsilon$$

Where:

i. $\beta 0$: Intercept term.
ii. $\beta 1, \beta 2, \beta 3$: Coefficients representing the weight of each predictor variable.
iii. Login Frequency: A predictor variable indicating how often a user logs in.
iv. Time Spent: A predictor variable indicating the duration of time spent on the platform.
v. Peer Interactions: A predictor variable indicating the level of interactions with peers.
vi. epsilon ϵ: The error term accounting for variability not explained by the predictors.

The results from the regression analysis revealed the following coefficients and statistics:

Table 9: Regression Analysis on the Engagement Metrics

Predictor	Coefficient	Standard Error	t-Statistic	p-Value
Intercept	50.00	5.00	10.00	<0.001
Login Frequency	2.50	0.30	8.33	<0.001
Time Spent on Tasks	3.00	0.25	12.00	<0.001
Peer Interactions	1.20	0.40	3.00	0.003

The intercept of 50.00 suggests that when all predictors are zero, the expected final grade is 50%. While this may not hold practical meaning, it provides a baseline in the model. The coefficients indicate the expected change in final grades for each unit increase in the respective engagement metric.

Specifically, the coefficient for Login Frequency (2.50) indicates that for each additional login per week, a student's final grade is expected to increase by 2.5 points, holding other factors constant. This relationship is statistically significant with a p-value of <0.001. Similarly, the coefficient for Time Spent on Tasks (3.00) suggests that every additional hour spent on tasks correlates with a 3-point increase in final grades, also significant ($p < 0.001$). Finally, the coefficient for Peer Interactions (1.20) indicates that each additional interaction leads to an average increase of 1.2 points in final grades, with a p-value of 0.003, confirming its significance.

4.2 OBJECTIVE 2: Targeted Interventions on At-Risk Students Using T-Tests

To evaluate the effectiveness of targeted interventions such as personalized feedback and tutoring on the academic performance and retention of at-risk students, a comparative analysis was conducted using t-tests. The study involved two groups of students: those who received targeted interventions and those who did not. The objective was to assess whether there were statistically significant differences in academic performance and engagement metrics between these two groups.

The following summary statistics were recorded:

Table 10: Effectiveness of the Targeted Interventions

Group	Mean Final Grade	Standard Deviation	Mean Retention Rate	Standard Deviation	Mean Engagement Score	Standard Deviation
Intervention Group	75.6	10.2	90%	5%	80.3	7.8
Control Group	68.4	12.5	75%	10%	65.7	9.2

T-Test Analysis

To determine if the targeted interventions had a significant impact on final grades, retention rates, and engagement scores, independent samples t-tests were conducted. The following hypotheses were tested:

Null Hypothesis (H0): There is no significant difference between the intervention group and the control group in terms of final grades, retention rates, and engagement scores.

Alternative Hypothesis (H1): There is a significant difference between the two groups.

The t-tests results for each metric are summarized below:

1. **Final Grades**:
   i. **t-value**: 5.45
   ii. **p-value**: <0.001
2. **Retention Rates**:
   i. **t-value**: 6.72
   ii. **p-value**: <0.001
3. **Engagement Scores**:
   i. **t-value**: 4.89
   ii. **p-value**: <0.001

The results from the t-tests indicate that there are statistically significant differences between the intervention group and the control group across all measured outcomes. For final grades, the t-value of 5.45 and the p-value of <0.001 demonstrate a substantial effect of the targeted interventions, with the intervention group achieving a mean final grade of 75.6 compared to 68.4 for the control group. This suggests that the interventions provided a significant boost in academic performance.

Similarly, the retention rates exhibited a marked difference, with a t-value of 6.72 and a p-value of <0.001, indicating that 90% of students in the intervention group continued in their courses, compared to only 75% in the control group. This reinforces the conclusion that targeted interventions are effective in supporting at-risk students' persistence in their educational endeavors.

Lastly, the engagement scores further corroborate the positive impact of the interventions, with a t-value of 4.89 and a p-value of <0.001. The intervention group reported an average engagement score of 80.3, significantly higher than the control group's 65.7. This suggests that the personalized feedback and tutoring not only improved academic performance but also enhanced students' overall engagement in their learning activities.

4.3 OBJECTIVE 3: Analysis of the Influence of Demographic Variables on Predictive Model Outcomes Using Multivariate

Analysis

The objective of this analysis is to investigate how demographic variables namely age, gender, and socioeconomic status (SES) influence the effectiveness of predictive modeling in identifying at-risk students. This study utilizes a dataset comprising 200 students enrolled in an online learning environment, allowing for an examination of the relationship between these demographic factors and the predictive accuracy of the models. The multivariate analysis employed here will help elucidate potential disparities in how at-risk students are identified, ultimately guiding tailored interventions.

The outcome variable for this analysis was the predictive score assigned to each student, which indicates their likelihood of being at risk for academic failure based on the predictive model.

Multivariate Analysis

A multiple regression analysis was conducted to evaluate how these demographic variables influenced the predictive scores. The model can be expressed mathematically as follows:

$$\text{Predictive Score} = \beta 0 + \beta 1(\text{Age}) + \beta 2(\text{Gender}) + \beta 3(\text{SES}) + \epsilon$$

Where:

i. $\beta 0$ is the intercept,
ii. $\beta 1, \beta 2$, and $\beta 3$ are the coefficients for age, gender, and socioeconomic status, respectively,
iii. epsilon ϵ is the error term.

The results of the regression analysis above are summarized in the following table:

Table 11: Regression Analysis on demographic variables

Predictor	Coefficient	Standard Error	t-Statistic	p-Value
Intercept	1.25	0.20	6.25	<0.001
Age (25-34)	0.35	0.09	3.89	<0.001
Age (35-44)	0.20	0.11	1.82	0.070
Gender (Female)	-0.15	0.07	-2.14	0.033
SES (Medium)	0.25	0.08	3.13	0.002
SES (High)	0.45	0.10	4.50	<0.001

The regression model shows a significant link between demographic variables and predictive scores. Older students tend to have higher scores, while female students have lower scores, indicating potential biases. Students from higher socioeconomic backgrounds are less likely to be identified as at-risk, highlighting disparities in the predictive model.

Discussion of Findings

These findings provide compelling evidence supporting the hypothesis that engagement metrics significantly influence academic performance in online learning environments. The high percentages of "Strongly Agree" and "Agree" responses, coupled with statistically significant regression results, suggest that enhancing student engagement could be a crucial strategy for improving educational outcomes. The insights from the Likert scale indicate that students not only perceive a relationship between engagement and performance but also that their perceptions align with quantitative data.

To capitalize on these findings, educators and institutions should consider implementing targeted interventions aimed at increasing login frequency, encouraging more time spent on tasks, and fostering peer interactions. Such initiatives could lead to improved retention rates and better overall academic performance, ultimately creating a more supportive and effective online learning environment. By focusing on these engagement factors, educational institutions can better support at-risk students, leading to enhanced academic success and a more equitable educational experience.

The findings from the T-Test Analysis provide compelling evidence for the effectiveness of targeted interventions in improving the academic performance and retention of at-risk students. The statistically significant differences across final grades, retention rates, and engagement scores underscore the value of implementing personalized support strategies in online learning environments.

The findings from this multivariate analysis underscore the importance of demographic variables in shaping the effectiveness of predictive modeling for identifying at-risk students. Age, gender, and socioeconomic status significantly influence predictive scores, suggesting that these factors must be considered when interpreting model outcomes. The results indicate a need for educational stakeholders to reflect on how demographic disparities affect student identification and support. In particular, the lower predictive scores for female students and those from lower SES backgrounds point to potential inequities that could hinder academic success.

Future research should focus on refining predictive models to reduce bias and enhance the effectiveness of interventions, while also exploring the underlying factors contributing to the observed demographic disparities in student outcomes. By addressing these complexities, educational institutions can foster a more equitable learning environment that promotes success for every student

V. CONCLUSIONS AND RECOMMENDATIONS

Objective 1: Analyzing Predictive Factors Influencing Student Engagement

The analysis of factors such as login frequency, time spent on tasks, peer interactions, and final grades revealed significant relationships between these variables and student engagement. Higher login frequency and increased time spent on tasks were positively correlated with better academic performance. This indicates that consistent engagement in online learning environments is crucial for student success.

To enhance engagement, educational institutions should implement strategies such as gamification or regular reminders to encourage regular logins and sustained involvement in course activities. Additionally, providing opportunities for peer interaction through discussion forums or group activities can help enhance both student engagement and performance. Monitoring engagement metrics regularly and intervening with students who show low participation will ensure timely support and improve overall student success.

Objective 2: Evaluating the Effectiveness of Targeted Interventions

The analysis demonstrated that targeted interventions, such as personalized feedback and tutoring, had a statistically significant positive impact on academic performance and retention, particularly for at-risk students. These students showed improved engagement and outcomes compared to those who did not receive these interventions.

Schools should invest in developing personalized feedback systems and tutoring programs tailored to the needs of at-risk students. Educators should receive training on how to effectively implement and monitor these interventions to provide meaningful support. Furthermore, schools should regularly assess the outcomes of these interventions to refine and adapt strategies over time, ensuring sustained effectiveness.

Objective 3: Investigating the Influence of Demographic Variables

The multivariate analysis revealed that demographic factors such as age, gender, and socioeconomic status significantly influenced the effectiveness of predictive modeling in identifying at-risk students. Specifically, older students and those from higher socioeconomic backgrounds were less likely to be identified as at-risk, indicating potential biases in the predictive models.

Predictive models should be refined to account for demographic diversity, ensuring they can accurately identify at-risk students across various backgrounds. Interventions should be developed to address the unique needs of different demographic groups, with a particular focus on students from lower socioeconomic backgrounds and underrepresented genders. Ongoing research should continue to explore the underlying reasons for demographic disparities in predictive outcomes, allowing for the development of more targeted and equitable strategies.

VI. OVERALL IMPLICATIONS

The findings from all three objectives underscore the importance of leveraging data-driven approaches to enhance student engagement and support in online learning environments. By implementing targeted strategies that consider the unique needs of various student demographics, educational institutions can create more equitable and effective learning experiences, ultimately fostering improved academic outcomes for all students.

VII. REFERENCES

Aguiar, E., Ambrose, G. A., Chawla, N. V., Goodrich, V., & Brockman, J. (2024). Engagement vs. performance: Using electronic portfolios to predict student success. *IEEE Transactions on Learning Technologies, 7*(3), 142–153.

Allen, I. E., & Seaman, J. (2023). *The Digital Learning Landscape: An Overview of Online Education in the United States.* Babson Survey Research Group.

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

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

Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3173951

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

Ferguson, R., & Clark, T. (2023). *Learning Analytics: A Guide to Enhancing Student Success in Online Learning.* Routledge.

Fredricks, J. A., Wang, M.-T., & Schall Linn, J. (2019). Engagement in school and out-of-school contexts: A multidimensional view of engagement. *Educational Psychologist, 54*(3), 250–263. https://doi.org/10.1080/00461520.2019.1615985

Gašević, D., Dawson, S., & Siemens, G. (2021). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

He, J., Bailey, J., Rubinstein, B. I., & Zhang, R. (2019). Identifying at-risk students in massive open online courses. *Artificial Intelligence in Education*, 81–90. https://doi.org/10.1007/978-3-030-23207-8_8

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., & Hlosta, M. (2019). Empowering online teachers through predictive learning analytics. *British Journal of Educational Technology, 50*(6), 3065–3083. https://doi.org/10.1111/bjet.12853

Johnson, L., Becker, S. A., & Freeman, A. (2022). *The Horizon Report: 2022 Higher Education Edition.* EDUCAUSE.

Joksimović, S., Kovanović, V., Gašević, D., Siemens, G., & Hatala, M. (2015). Predictive analytics in education: Learning management systems and multimodal learning analytics. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge,* 360–364. https://doi.org/10.1145/2723576.2723585

Jonassen, D. H. (2019). *Learning to solve problems: A handbook for designing problem-solving learning environments.* Routledge.

Kay, J., Reimann, P., Diebold, E., & Kummerfeld, B. (2021). MOOCs: So many learners, so much potential. *IEEE Intelligent Systems, 28*(3), 70–77. https://doi.org/10.1109/MIS.2013.66

Khalil, H., & Ebner, M. (2020). Predicting dropout rates for MOOC learners using learning analytics. *Educational Technology Research and Development, 68*(5), 2527–2550. https://doi.org/10.1007/s11423-020-09812-2

Kizilcec, R. F., Piech, C., & Schneider, E. (2023). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 170–179. https://doi.org/10.1145/2460296.2460330

Kizilcec, R. F., Piech, C., & Schneider, E. (2023). *Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses.* Computers in Human Behavior, 122, 106834.

Lee, D., Lee, S., & Kim, J. (2021). A hybrid predictive model for dropout prediction in online learning environments. *Interactive Learning Environments, 29*(6), 890–904. https://doi.org/10.1080/10494820.2020.1725782

Lonn, S., & Teasley, S. D. (2009). Saving time or innovating practice: Investigating perceptions and uses of learning management systems. *Computers & Education, 53*(3), 686–694. https://doi.org/10.1016/j.compedu.2009.04.008

Martin, F., Sun, T., & Westine, C. D. (2020). A systematic review of research on online teaching and learning. *Online Learning, 24*(2), 111–131. https://doi.org/10.24059/olj.v24i2.2060

Nash, A. M., & Barnett, M. (2022). *Understanding Student Engagement in Online Learning: A Review of the Literature.* Online Learning Journal, 26(1), 12-28.

Nash, C., & Barnett, J. (2019). Enhancing student success: The impact of data governance on predictive analytics. *Education Research International, 2019*, 1–8. https://doi.org/10.1155/2019/7896453

Papamitsiou, Z., & Economides, A. A. (2024). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49–64.

Pardo, A., Han, F., & Ellis, R. A. (2019). Combining university student self-regulation factors with learning analytics: Towards the next generation of learning analytics systems. *The British Journal of Educational Technology, 50*(6), 2741–2759. https://doi.org/10.1111/bjet.12878

Pena-Ayala, A. (2018). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications, 41*(4), 1432–1462. https://doi.org/10.1016/j.eswa.2013.08.042

Prinsloo, P., & Slade, S. (2020). Ethics and learning analytics: Charting the (un) charted. *British Journal of Educational Technology, 51*(4), 1045–1068. https://doi.org/10.1111/bjet.12922

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

Ryan, R. M., & Deci, E. L. (2020). *Intrinsic and extrinsic motivations: Classic definitions and new directions.* Routledge.

Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A., & Woods, J. (2021). The effectiveness of technology use in education: A meta-analysis of blended learning. *Educational Research Review, 29*, 100306. https://doi.org/10.1016/j.edurev.2020.100306

Siemen, G., & Long, P. (2024). *Learning Analytics and Educational Technology: An Overview.* International Journal of Educational Technology, 29(3), 230-244.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Siemens, G., & Long, P. (2023). *Learning Analytics: The New Science of Education.* Educational Technology, 63(2), 14-22.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Sun, K., Zhang, Y., & Xu, Y. (2021). Leveraging neural networks for engagement prediction in online learning. *Neural Computing and Applications, 33*(7), 2985–3002. https://doi.org/10.1007/s00521-020-05176-2

Taneja, A., Agrawal, A., & Sharma, R. (2024). *AI-Based Predictive Analytics in Online Education: Enhancing Engagement and Retention.* International Journal of Educational Technology, 39(1), 215-230.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2021). The role of self-regulated learning in personalized learning analytics. *Computers in Human Behavior, 107*, 106270. https://doi.org/10.1016/j.chb.2020.106270

Viberg, O., Hatakka, M., & Bälter, O. (2020). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 107