Journal of Higher Education Policy and Management

ISSN: 1360-080X (Print) 1469-9508 (Online) Journal homepage: www.tandfonline.com/journals/cjhe20

Strategic use of predictive analytics for student retention in blended higher education: implications for education management

İrfan Ayhan

JOURNAL OF HIGHER EDUCATION POLICY AND MANAGEMENT
https://doi.org/10.1080/1360080X.2026.2643372

Strategic use of predictive analytics for student retention in blended higher education: implications for education management

İrfan Ayhan

Kaman School of Applied Sciences, Kırşehir Ahi Evran University, Kırşehir, Türkiye

ABSTRACT

This research investigates how predictive analytics can support student retention in blended higher education. Using the engagement and assessment data of 523 students from 3 universities, four machine learning models were developed. Of these four models, Multilayer Perceptron achieved the highest accuracy (70.0 per cent) and after handling class imbalance performed better at identifying students that are at risk of failure. Utilising theories of self-regulated learning, behavioural engagement, and academic integration, the study identifies the indicators that are most frequently associated with successful outcomes. Predictions of machine learning models were compared to the early predictions of instructors and revealed different error patterns, suggesting that models and instructors capture different dimensions of student engagement. Most significant predictors were learning management system activity, pacing, assessment behaviours, and attendance. Findings suggest that predictive analytics can support early warning systems as well as guiding institutional strategies and strengthening ethically responsible retention strategies in blended higher education.

ARTICLE HISTORY

Received 2 August 2025
Accepted 6 March 2026

KEYWORDS

Predictive analytics; educational management; student retention; blended learning; learning analytics; early intervention

Introduction

As student attrition rates continue to increase and institutions are held accountable for their activities, higher education leaders have a major challenge to promote academic success and retention of students in digitally mediated environments. Blended learning models are considered an educationally innovative approach; however, they pose additional challenges to provide adequate oversight of student engagement on a large-scale. Therefore, there is an increasing urgency to provide educational managers and academic leaders with reliable, interpretable, and actionable data-driven tools to facilitate timely decision-making and the allocation of resources. Technology continues to be integrated into higher education; therefore, blended learning – which combines traditional classroom instruction and online content – will continue to increase. By using this model, students have the opportunity to utilise their learning experiences more flexibly, whereas educators can monitor and evaluate their students' learning performance through the use

CONTACT İrfan Ayhan irfanayhan@gmail.com Kaman School of Applied Sciences, Kırşehir Ahi Evran University, Hacıpınar Mah. Ş. Uzm. Ç. Bahtiyar Şimşek Cad. No: 147 Kaman, Kırşehir, Türkiye
Supplemental data for this article can be accessed online at https://doi.org/10.1080/1360080X.2026.2643372

of Learning Management Systems (LMS) (Anthony et al., 2020; Nguyen & Dao, 2024). When examining the data collected through LMS, educators can have access to a wide range of students' behavioural indicators, including login frequency, forum participation, time on task and assignment and exam results. Each of these indicators has the potential to provide educators with the knowledge and understanding of how to improve and optimise their students' learning experiences.

In this context, the field of learning analytics has gained prominence as a methodological framework for leveraging educational data to improve teaching and learning processes (Guzmán-Valenzuela et al., 2021). Specifically, learning analytics involves collecting and analysing data about students to identify patterns that can help guide decisions and improve educational outcomes (Siemens, 2013). By using machine learning, learning analytics can move beyond just describing data to predicting which students might be at risk of falling behind or disengaging (Kok et al., 2024).

In blended learning, being able to predict a student's likelihood of success is extremely important when an instructor has less visibility on a student's level of engagement due to the difficulty of determining the level of engagement through observations compared to their actual work. By utilising historical data, machine learning models such as logistic regression and decision trees can be trained, and these models can then be used to determine the complex relationships among many different types of student behaviours and their academic performance (Tiwari & Jain, 2024). Although these models still need more testing, they can provide valuable insights to help create systems that identify students at risk and offer personalised support (Almalawi et al., 2024).

Predictive analytics provide many useful insights; however, they also bring ethical and moral challenges for education, including privacy, fairness, transparency, and consent from students (Guan et al., 2023; Mishara, 2024; Slade & Prinsloo, 2013). Given these ethical concerns, institutions should provide established guidelines for the appropriate application of predictive analytics and establish equitable practices governing access to and use of student data (Gandara et al., 2024).

A large number of studies about learning analytics show that LMS-based behavioural indicators (quizzes completed, number of posts in forums, time spent completing tasks) can predict student performance (Ifenthaler & Yau, 2020; Waheed et al., 2020). However, most of these studies focus on the technical aspect of these models and have not provided much insight into the ways in which these models might be interpreted for the classroom (Guzmán-Valenzuela et al., 2021). Even though Random Forest and XGBoost models are successful at predicting student performance (Tiwari & Jain, 2024), few studies have compared their predictions with those of instructors when making decisions regarding student performance. Research on instructor judgement shows that early predictions are moderately accurate but often shaped by limited behavioural cues and biases (Karademir Coşkun & Alper, 2024; Shavelson et al., 2008), making it essential to explore how machine predictions can complement, rather than replace, educators' expertise.

While predictive analytics and learning dashboards are a growing area of interest, most of the available literature on educational management research is descriptive in nature. Existing research tends to be retrospective rather than focusing on predictive tools for real time decision making and intervention-based strategies (Ifenthaler & Yau, 2020; Leitner et al., 2017; Tsai et al., 2019). Current educational management research has been limited in considering how to create action-oriented retention strategies, advising, and student success initiatives from a predictive perspective (Macfadyen & Dawson, 2012; Paolucci et al., 2024; Zawacki-Richter et al., 2019). To address this gap in the literature, the present study provides a theoretical basis for developing a predictive framework that is empirically validated and has implications for policy, resource allocation, and early-warning systems within higher education.

Such challenges to higher education systems are well documented but are also affected by the unique nature of national policy contexts and institutional settings. A national higher education system has its own context for examining issues of engagement, risk, and retention. As such, the Turkish context provides a rich opportunity to better understand how these factors play out in practice. The Turkish higher education system's unique environment is characterised by a strong centralised bureaucratic structure, rapid growth in the number of blended learning programmes available, and increased attention to using data and evidence as the basis of quality assurance measures for higher education.

Student retention patterns in Türkiye vary widely by level of analysis and metric. Based on OECD data, approximately 1 per cent of first-time entrants into bachelor's programmes are dropping out after their first year (OECD, 2025). This indicates student persistence at a system level rather than disengagement from the course. However, national completion and administrative data indicate more widespread and complicated long-term retention patterns, including delayed graduation and cumulative attrition rates along the degree completion trajectory (YÖKAK, 2025). Importantly, research examining technology-mediated and blended learning environments in higher education indicates higher course-level withdrawal and failure rates than what existing drop-out indicators estimate (Elibol & Bozkurt, 2025; İlkdoğan Serbes & Balyer, 2025; Usta Ergün, 2025). Consistent with these findings, the percentage of students identified as academically at risk in this study's blended-learning sample (~30 per cent) was much higher than national estimates of first-year dropouts. This difference reinforces the importance of analysing student risk and persistence at the course and engagement level, particularly in blended learning environments where formal enrolment may mask patterns of early student disengagement.

There are several theories that have looked at student retention, each viewing student persistence and withdrawal as resulting from several different mechanisms that support and operate together rather than in isolation. According to Tinto's Theory of Student Departure (Tinto, 1993), students who are both academically and socially integrated into their institution are more likely to persist than those who are not. Tinto defines academic integration in terms of the extent to which a student is engaged with learning activities and performs well academically, while social integration is defined in terms of how much a student interacts with their peers and instructors. If either area becomes disengaged, the likelihood of attrition will increase. Self-Regulated learning theory (Zimmerman, 2002) views student persistence as a process of planning, monitoring, and regulating cognitive and behavioural strategies. There are many ways in which this process can be observed in the context of blended learning. For example, students who are actively participating in blended learning are often more likely to engage with their learning materials on a regular basis, turn in assessments in a timely manner, and interact with their classmates and instructors regularly. Constructivist learning theory (Jonassen, 1994; Vygotsky, 1978) expands upon the concept of persistence as a socially constructed process whereby knowledge is constructed within the context of a community of learners through interaction and dialogue, therefore collaborating and engaging in conversations about the material being studied are key methods to support persistence in the context of technology-enhanced learning. Behavioural engagement theory (Fredricks et al., 2004) expands on the previous three constructs and focuses specifically on the observable behaviours of students (i.e., participation, effort, and task completion). The behavioural engagement theory gives a heuristic framework to the study of student engagement; by using data from student behaviour (trace data), researchers can see what level of student engagement is sustained or has become disengaged over time.

Building on these complementary perspectives, engagement indicators in this study were operationalised using LMS trace data that reflect theoretically meaningful academic, behavioural, self-regulatory, and social processes. The measures of academic and behavioural integration of students are encapsulated by the level of integration among academic and behavioural elements of participation. Academic integration measures include attendance, the number of quizzes taken, and the frequency of accessing course materials (Tinto, 1993), while behavioural integration is represented by observable student effort and participation in terms of time spent on course activities as indicated by students' individual logins, pacing, completion of assignments on time, persistence and other self-regulated learning behaviours (Fredricks et al., 2004; Zimmerman, 2002). These measures together comprise a multi-faceted representation of academic and behavioural integration in a blended learning context (e.g., Jonassen, 1994; Vygotsky, 1978). This research examines all student engagement behaviours as a collection of indicators from multiple theoretical frameworks, rather than assigning indicators to singular theoretical categories. The map created from this theoretical analysis is presented in Appendix A (online supplementary materials).

In addition to behavioural engagement, recent research also demonstrates the importance of motivational and emotional dimensions in blended learning. Intrinsic motivation and emotional engagement have been positively correlated with higher levels of performance (Liu et al., 2024). Autonomy has been determined to be a predictor of emotional engagement (Li & Zeng, 2025). While motivation and emotion have long been viewed as significant to understanding engagement in a blended form of learning, the present research seeks to examine the behavioural and interactional aspects of engagement in blended learning, as they can more easily be obtained from data collected by the LMS.

Using these theoretical perspectives allows the study to fill gaps in previous studies of learning analytics by creating interpretable predictive models, comparing the results of predictive modelling with teacher assessments, and evaluating how these findings can benefit classroom practice while ensuring ethical responsibilities. Guided by this approach, the study explores the following research questions:

- RQ1: What factors influence student engagement and are actionable to support students' persistence in blended learning environments?
- RQ2: How do the predictions made by machine learning systems compare with the assessments provided by instructors from a management perspective?
- RQ3: What ethical and operational considerations must be taken into account when implementing predictive analytics within educational management systems?

- RQ4: How can predictive analytics be used to develop scalable interventions to support students throughout their learning experiences?

Methodology

Context of the study and blended learning implementation

The present research collected data from three institutions of higher education in Türkiye; a large public university and two medium sized foundation universities granting bachelor's degrees in Computer Science, Business and Education through the use of blended learning methods. All three of these institutions use a variety of learning management systems including Moodle or Blackboard and blended learning was chosen in this case as it generates an ongoing supply of structured LMS trace data based on how students interact within that environment.

In combination with weekly classroom meetings for instruction, students also access asynchronous content, complete quizzes and post to discussion forums. In addition, instructors may schedule synchronous online meetings periodically. Through an LMS, all student interaction with these blended learning components is captured and this was the primary source of data for analyses in this research. Therefore, the multi-institutional blended learning model of higher education provides insight to the 'operationalisation' and to the challenges faced by educational leaders. Since the data generated through the LMS will continuously provide insight into student activity and engagement, higher education managers must assess these engagement data in aggregate to help shape retention policies and resource allocation decisions.

Study design and data collection

This research utilised a quantitative research method to predict the academic achievements of students enrolled in a blended learning environment through the use of learning analytics and machine learning. The main goal of this research was to discover ways in which predictive modelling could shed light on teaching behaviours and create early intervention tools for ensuring success among students. The sample consisted of 523 students from three areas of study – Computer Science ($n = 198$), Business ($n = 171$) and Education ($n = 154$) – whose data was extracted from the institutions' learning management systems. Key indicators of engagement factors such as frequency of logins, total time spent using the LMS, number of quizzes completed, number of forums posted to, and number of assignments submitted were recorded; demographic and programme variables (age, gender, area of study, and year level) were also included for analysis of differences across the three groups on predictive performance (as shown in Table 5). Two types of predictive models were developed-one model that was based solely on the LMS behaviour and one model that included a combination of behaviour as well as demographic variables. Demographic information was only included in the analyses and not used in making decisions through automation.

Participation was voluntary, and only anonymised data from students who provided written informed consent was included in this research. Due to privacy regulations, the number of students who did not give consent to use their data was not revealed; however, instructors reported that there were no significant differences in the performance of consenting and not consenting students. Categorical variables were coded, and continuous variables were standardised for consistency. For the purpose of evaluating model performance across the three groups of students, subgroup analyses by discipline and year level were conducted to identify areas that warranted additional scrutiny.

The data collected included both engagement and course performance data over the course of a semester. The selection of predictive features was based on theory; measures of engagement were selected based on theoretical constructs as summarised in Table A1 in Appendix A. All data were anonymised to ensure the protection of student privacy; informed consent was obtained from all students prior to the inclusion of their data. The ethical handling of sensitive student data allows for valuable insights into student engagement and student academic performance. Additionally, the evaluation of LMS data provides insights for institutional leaders about the capacity to use routinely collected data to identify students at risk of academic failure and enable them to monitor and assess the effectiveness of their retention efforts without adding to the burden of reporting.

Data preprocessing

The data were cleaned before being analysed, and missing values were imputed based on the average value. Participation levels (a categorical classification) were converted to their numerical equivalent so they would work with machine learning models, and all data were standardised to ensure consistency. To counteract the class imbalance in the data (70.0 per cent passing, 30.0 per cent failing) the Synthetic Minority Over-sampling Technique (SMOTE) was used to generate synthetic (artificial) observations for the under-represented group (students who were failing). Using this method enables the model to predict which students are likely to be at risk without having an inherent bias towards the majority class. While the use of SMOTE increases the accuracy of the model, it is possible that some noise may be introduced because the synthetic observations do not necessarily represent actual student behaviour. This trade-off is acceptable in order to facilitate early intervention with at-risk students (Althaqafi et al., 2025; Law et al., 2024). Education managers also need to address the issue of the class imbalance in order to ensure that their retention systems identify those students who may need additional support; even if this causes a minor decrease in overall accuracy, retaining these students will prevent the inadvertent neglect of at-risk populations.

Model development and evaluation

This study employed four different machine learning methods to predict how well students would perform: Logistic Regression, Random Forest, XGBoost, and Multilayer Perceptron (MLP). The reason for selecting these methods was because they all offer different types of classification capabilities, providing a very thorough way to evaluate the ability of each of these algorithms to predict student outcomes based on the information available at the time of prediction. Logistic Regression is often considered to be a simple model that is easy to interpret, while Random Forest and XGBoost are both examples of ensemble models. Ensemble models tend to be more robust and allow for better treatment of non-linear relationships within the data. Being a deep learning model, MLP is very powerful at capturing complex relationships and patterns in data. These four types of models have an extensive range of complexity; consequently, using all four provides a variety of viewpoints with respect to the effectiveness of each type of model for educational data. The effectiveness of using these machine learning models to assist educators with student success is supported with research conducted by authors such as Berens et al. (2025), who highlighted the widespread use of Random Forest and other models for performance prediction and dropout detection; Carballo-Mendivil et al. (2025), who noted that ensemble methods like XGBoost provide improved predictive power for educational interventions; and Rebelo Marcolino et al. (2025), who includes a cluster-based approach to analyse dropout patterns using LMS data.

The study investigated the effectiveness of models that predict if a student will pass or fail a specific course with the risk of failing indicating potential academic risk. To evaluate how well the models performed, the researcher split the data, using 80 per cent for training and 20 per cent for testing. Several metrics, including accuracy, precision, recall, and F1-score, were utilised to assess the predictive power of the models. Accuracy is defined as the percentage of correct predictions made by a model and is used as an overall indication of the model's predictive power (Sokolova & Lapalme, 2009). In an educational setting, however, solely relying on accuracy for predictive evaluations could be misleading, given that the majority of at-risk students represent a very small segment of the student population.

Recall measures how well the models identify at-risk students who really could fail. From an administrative perspective, recall is very important; without an adequate recall rate, the students needing assistance will go undetected and therefore, miss out on timely assistance. Precision measures what percentage of students identified by the models as being at risk are actually at-risk. It indicates how support is efficiently allocated to students who are truly at risk of failing and could help determine the potential cost associated with unnecessary interventions. The F1 score (Powers, 2011) is a calculated value of precision and recall; the F1 score represents a balanced evaluation of model performance in real life, where missed students or unnecessary interventions have practical implications for how managers make decisions related to institutional resources. If education managers choose to value recall over accuracy, the models will assist them in fulfilling their organisational goals related to student retention since missed at-risk students represent a lost opportunity to provide support in a timely manner and optimise the use of support resources.

Cross-validation was applied by the researcher to verify that the final model had a high degree of accuracy and was able to provide the same results across different data sets. The purpose of cross-validation was to reduce the likelihood of the final model being trained on a single sample, thereby increasing its potential to successfully make predictions about new data sources – something especially critical when developing predictive models specific to education (Guevara-Reyes et al., 2025). Though cross-validation can greatly enhance the reliability of a model, it can also consume a considerable amount of time and processing power, particularly for very large and complicated data sets.

In addition, the model's predictions were compared to those of instructors after four weeks based on their own observations, which allows for the evaluation of how effectively machine predictions can supplement instructor's assessments. Through the use of statistical testing, the researcher assessed the extent of the similarity between the instructors' and the machine's predictions. To make the models easier to interpret and more usable by instructors, the researcher utilised two explainable artificial intelligence tools (SHAP and LIME) to assist in determining which factors (e.g., quiz scores and student engagement) have the greatest impact on predicting which students would be considered 'at risk' by the model and to assist instructors in understanding the rationale behind the models' decisions. While SHAP was able to provide insights into general trends and patterns among students, LIME focused on specific predictions for each student. The results of this comparison are particularly valuable to educational administration, as they illustrate that predictive analytics can be implemented as a decision-support layer that helps inform professional judgement rather than replacing traditional advising and instructional practices.

Ethical considerations regarding data privacy were adhered to in this study through the anonymisation of all student data as well as the secure storage of that information. The study complied with the institutional review board policies, as well as The General Data Protection Regulation. Appendix B of the online supplementary materials contains the Python code that was used to conduct the data analysis, including the processes of pre-processing, training and evaluating the models. This code was run in a Google Colab environment. The selection of the models is designed to create an appropriate balance between being interpretable, predictive and usable as an operational tool (for providing managerial oversight as opposed to automating) in order to allow for informed decision-making by managers.

Findings

Of the 523 students that were studied, 369 (70.6 per cent) passed their classes while 154 students (29.4 per cent) were identified as being at risk of failure. Academic risk in this study is defined by identifying students who are at increased risk of non-completion or delayed progression based on their level of engagement and assessment performance in their courses; this does not consider whether or not students withdrew from an institution. The rates of failure observed herein are similar to those reported in the literature for blended and technology-enhanced learning environments, with at-risk students having significantly higher rates of failure than the national average of 1st-year college students who drop out (Deschacht & Goeman, 2015; Millimouno et al., 2022; Rahmani et al., 2024). The imbalance in the observed class sizes (approximately 70.0 per cent students passing and 30.0 per cent failing) impacted the modelling that followed and led to the use of resampling methods to better identify students who were at risk. The main predictors included: frequency of logins; total time spent in the LMS; quiz scores; forum posts; and attendance at class (both online and in person). Table 1 lists the descriptive statistics of each of these predictors:

Table 1. Descriptive statistics for student dataset.

Variable	Mean	Std. Dev.	Min	Max
Logins	24.8	14.5	0	49
Total Time Spent (min)	255.4	144.7	10	499
Quiz Scores (%)	70.6	17.2	40	99
Forum Posts	9.6	5.9	0	19
Attendance (%)	74.8	13.4	50	99

Four machine learning models were trained and evaluated: Logistic Regression, Random Forest, XGBoost, and Multilayer Perceptron (MLP). These models were evaluated based on their accuracy in predicting student success (pass/fail). The results, summarised in Table 2 below, show that the MLP model performed the best, with an accuracy of 70.0 per cent. Therefore, it was selected for further analysis due to its superior performance.

Table 2. Model accuracy (cross-validation).

Model	Accuracy (%)
Logistic Regression	68.0
Random Forest	69.0
XGBoost	68.5
MLP	**70.0**

When the model was tested on an independent dataset its overall accuracy decreased (compared with previous tests), yet cross-validation revealed that the model's accuracy was consistent across all tests at a level of 70.0 per cent. The ability of the model to accurately identify students at risk of failing was an essential factor when considering prevention methods for those students, as early intervention is critical to a student's success. The MLP model demonstrated high accuracy with successful students but provided less precision with unsuccessful students; this is typically due to the problem of unbalanced data. By using SMOTE to artificially create additional data for the failing students, the data became balanced resulting in improved accuracy of the model to identify students at risk of failing. After implementing the SMOTE technique, improvements were achieved for recall as well, which increased to 0.68 for the students who were unsuccessful in completing the course, indicating a much higher accuracy in identifying students in need of assistance. See Table 3 for summary of findings:

Table 3. Classification report of the MLP model (Post-SMOTE).

Class	Precision	Recall	F1-Score	Support
Fail (0)	0.60	**0.68**	**0.64**	34
Pass (1)	0.75	0.72	0.73	66
Overall	–	–	**0.67**	100

Also, machine-generated predictions were compared to those made by instructors, who predicted student outcomes after four weeks based on early engagement data from the LMS. The comparison showed that while the MLP model had slightly higher accuracy (70.0 per cent vs. 69.4 per cent for instructors), instructors had a higher Fl-score. This means that while the MLP model was slightly more accurate in predicting overall student success, the instructors were better at correctly identifying both successful and failing students. Table 4 below summarises the accuracy, Fl score, and match rate with actual student outcomes:

Table 4. Instructor vs model prediction performance.

Metric	Instructor Prediction	MLP Model
Accuracy	**0.69**	**0.70**
F1-Score	**0.70**	0.67
Match Rate with True Labels	0.61	**0.62**

By demonstrating the potential of the MLP model, these findings suggest that MLP modelling is capable of providing insights beyond those offered by an instructor's limited assessment of overall student performance. While instructors may have a better sense of the students' general performance level, the MLP can excel in identifying which students may be at risk of failure (highly important for early intervention strategies) and thus provide an additional means of supporting the instructor in a timely manner. The recall value of 0.68 for at-risk students demonstrates that the MLP could offer sufficient reliability to serve as an early warning system for educators to determine which students require the most attention in terms of academic support. To gain more insight into how the MLP identifies at-risk students, SHAP and LIME were used. The four most important features influencing the MLP's predictions were total time spent on the LMS, quiz scores, logins, and attendance.

The SHAP and LIME analyses provided further clarification regarding these features' relationship to dropout risk. Students who spent limited time on the platform or logged very few times into the LMS were very likely to fail. Students who exhibited moderate to well-paced levels of LMS engagement were most frequently associated with positive outcomes. Students who performed well in quizzes (typically achieving a mean of more than 85 per cent) were consistently associated with a high probability of passing, while students with poor quiz performance were more likely to fail. It is important to note that success is not linear and both SHAP and LIME provided similar conclusions regarding the predictive power of time-on-task when measuring student success; that is, both indicated that spending too much or very little time in the LMS is inversely related to success. Attendance had a positive relationship to student success, however, attendance as a predictor of success can be viewed as a secondary predictor when compared with ongoing LMS behaviours and performance in quizzes. The overall takeaway from these findings is that quality engagement is more important (as an indicator) than how much activity is conducted in the LMS in predicting dropout risk among blended learners. Table A2 in Appendix A gives a list of the feature importance rankings from the SHAP analysis and Table A3 shows a detailed explanation of LIME results of individual students.

The results indicate that the model is largely consistent across all student cohorts, but the level of consistency varies depending on which student group is evaluated. Overall, Computer Science students and students in their later years of study obtained the most consistent model predictions; this is likely because they generally demonstrated more consistent student engagement behaviours in the LMS than those in other programmes, such as Education, where learners displayed much greater variability. As the MLP model was able to more accurately identify the majority of at-risk students in all groups when demographic and programme variables were included, the difference between groups decreased. Compared to the MLP model with only student behaviour, the complete MLP model (including demographic control variables) demonstrated consistent but small increases in predictions of at-risk students in accuracy, F1, and recall. Table A4 in Appendix A summarises the results for behaviour-only and the complete model. There were no subgroups who were consistently false-positive/false-negative classified and the overall accuracy remained relatively close to the first model. The results are presented below (Table 5):

Table 5. Subgroup performance metrics (MLP model after SMOTE adjustment).

Subgroup	N	Accuracy	Precision	Recall (Failing Students)	F1-score
Gender					
Male	264	0.71	0.67	0.69	0.68
Female	259	0.69	0.65	0.66	0.66
Field of Study					
Computer Science	198	0.73	0.69	0.70	0.69
Business	171	0.69	0.65	0.67	0.66
Education	154	0.67	0.63	0.65	0.64
Year of Study					
1st year	181	0.66	0.62	0.71	0.66
2nd year	170	0.69	0.65	0.67	0.66
3rd year	110	0.72	0.68	0.66	0.67
4th year	62	0.74	0.70	0.64	0.67
Age Groups					
≤20	242	0.67	0.63	0.69	0.66
21–24	264	0.71	0.67	0.67	0.67
≥25	17	0.69	0.66	0.64	0.65
Overall	523	**0.70**	**0.66**	**0.68**	**0.67**

These findings highlight how useful predictive models can be in higher education with regard to helping institutions prioritise intervention and retention strategies. The use of machine learning models on institutional dashboards gives educational leaders access to real-time data that allow them to identify students who are at-risk and provide targeted support. Explainable AI tools provide transparency, which allows for instructors and decision makers to utilise predictive models to identify students who may need additional assistance by accompanying the predictive model data with existing instructor assessments. Additionally, the predictive data utilised in this study enhances institutions' ability to allocate support resources and ensures that services are more effectively delivered to students who require assistance.

Discussion

Implications for retention policy

The findings demonstrate that patterns of continuous engagement and assessment behaviours in learning management systems are the best predictors of academic risk as compared to static measures, such as attendance. Therefore, education leaders need to keep in mind the importance of tracking engagement in real-time rather than relying on retrospective measures such as attendance. Institutions that place a greater emphasis on consistent engagement and pacing, as opposed to simply volume, will identify at-risk students sooner and allocate resources more efficiently, leading to timely and appropriately targeted interventions (Fan et al., 2024). Importantly, the relationship between LMS engagement and success is non-linear, both very low and high levels of LMS engagement are associated with lower academic success.

There are several theoretical perspectives to explain this pattern. Self-regulated learning theory argues that excessively high levels of LMS activity demonstrate maladaptive self-regulation where students do not monitor their own progress or are unable to set appropriate goals (Zimmerman, 2002). Cognitive Load Theory suggests that prolonged participation in a learning management system can create cognitive overload that precludes an increase in academic success (Sweller et al., 2011), while behavioural engagement theory indicates that having a higher quantity of participation does not necessarily indicate higher quality (Fredricks et al., 2004). These insights are consistent with the multiple dimensions of engagement in blended learning that include cognitive, motivational/emotional, and behavioural components (De Bruijn-Smolders & Prinsen, 2024; Liu et al., 2024). Therefore, institutions have to build predictive tools that measure not only the engagement but also the quality of engagement to allow for informed decision-making on the implementation of retention strategies (Goh, 2025). These findings contribute to the development of early-warning systems, which can assist institutions to implement early and timely interventions (Macfadyen & Dawson, 2012).

Model utility for educational leaders

From an educational administration standpoint, the predictive model's value is not only derived from increasing the accuracy of predictions, but also providing educational administrators with insights into large scale student engagement behaviour. Instead of using traditional measures such as attendance, the model uses more dynamic student engagement signals. The use of a dynamic student engagement signal aligns with Tinto's theory (1993), which posits that long-term commitment to learning is necessary for student persistence. The model's ability to address issues of class imbalance enables it to have a greater sensitivity to at-risk students and allows administrators to provide timely interventions based on the prioritisation of student support over accuracy alone. The need for an information-based approach to provide early interventions based on formulative assessments is supported by a growing body of research linking successful interventions for high-risk learners with the timing of those interventions (Chang et al., 2025; Santos & Henriques, 2023; Susnjak et al., 2022). For educational managers, this means using the model to allocate resources in a manner that ensures early identification and immediate support for at-risk learners throughout their academic journey.

Additionally, in the light of comparative results of predicting student outcomes through machine learning and instructors' assessments of students, this model demonstrates the role of predictive models to support human intuition. While instructors can frequently and accurately assess students overall, predictive analytics are capable of identifying specific risk factors, therefore can assist educational managers with prioritising student interventions. This establishes predictive models as decision-support tools, rather than as substitutes for instructors' judgement (Ninaus & Sailer, 2022). The model demonstrates a recall of 0.68 for at-risk students, thereby emphasising its ability to identify students requiring support and allowing proactive measures to be taken.

The predictive performance of the model's behaviour data is approximately 70.0 per cent, which aligns with other studies that have used similar machine learning approaches to produce early warning systems in the higher education context. Similar machine learning research demonstrates that LMS-based behavioural indicators, such as time spent on tasks, engagement with quizzes, and other learning activity patterns, serve as effective predictors of student success (Carballo-Mendivil et al., 2025; Villar & de Andrade, 2024). The model also exhibited cross-validation stability, providing further evidence that engagement mechanisms such as academic integration (Tinto, 1993) and self-regulation (Zimmerman, 2002) accurately predict student success across cohorts. As such, the model provides education leaders an opportunity to improve retention across diverse student populations and within a variety of institutional contexts.

Ethical oversight and contextual interpretation

In terms of ethical oversight and contextual use, there are many ethical considerations in using predictive analytics in higher education, including potential bias, performance transparency, and appropriate use of data. Because predictive models rely on historical data to train their algorithms, using those outputs without accounting for historical bias could implement new barriers and reinforce existing inequities for one or many groups of students who may be disadvantaged as a result (Gandara et al., 2024). One way to promote responsible use of predictive analytics is to utilise explainable AI tools, such as SHAP and LIME, to promote transparency by demonstrating how data collected through engagement-related activities leads to classifications of risk. Using explainable AI can help build trust between stakeholders and encourage responsible use (Holstein et al., 2019). Ethical guidelines related to predictive analytics recommend that predictive outputs should be used to support intervention rather than punitive action, and demographic data should be used to inform the interpretation and fairness audits of these results, but not used in an automated manner to make student-related decisions (Khosravi et al., 2022; Slade & Prinsloo, 2013).

Subgroup analyses indicated differing predictive performance across disciplines and year levels, with first-year students being much less predictable than more experienced students due to their variable engagement patterns. For students who are early in their programme, they may have yet to establish consistent strategies for planning and monitoring their learning, making the predictions less reliable (Zimmerman, 2002). Moreover, different uses of the LMS – such as task-based activities in Computer Science versus discussion-based engagement in Education – show the importance of being sensitive to context for interpreting behavioural data (Opazo et al., 2021; Whitelock-Wainwright et al., 2020). This provides support for the rationale of tailoring intervention strategies to the context of a given programme and stage of student; thus, risk alerts should be regarded as contextual indicators rather than universal thresholds for decision-making.

Translating predictive insights into actionable interventions

Implementing effective interventions requires addressing the mechanisms behind disengagement and not merely addressing the risk flags indicated by predictive systems. Self-Regulated Learning (Zimmerman, 2002) and constructivist theories (Pratama et al., 2025) highlight the significance of providing support for students' planning and monitoring behaviours that promote engagement. Therefore, if aligned with these theories, predictive analytics would assure that early warning systems facilitate the delivery of targeted and pedagogically grounded interventions such as academic advising, tutoring, and community-building (Hawthorne et al., 2021). For educational leaders, this means using predictions within institutional systems to provide timely support to improve student engagement and retention.

Effective use of predictive models also depends on the degree to which they can be trusted, are interpretable, and have the potential to be replicated across situations. Predictive models can be integrated into LMS dashboards and can provide teachers and staff with automated alerts, recommendations, and visual explanations via SHAP and LIME to support them while delivering the course (Holstein et al., 2019). For these tools to achieve successful use, it is essential that they are pedagogically relevant, understandable, and aligned with the organisation's current workflows (Khosravi et al., 2022). Standard LMS indicators, such as logins and forum contributions, that are used in predictive models will make the models transferrable across diverse institutions and blended learning environments provided that local pedagogical practices and technological environments are accounted for in their use. Therefore, predictive analytics can serve as a scalable and context sensitive tool for making decisions related to retention strategies and supporting leadership decisions in higher education.

Limitations

The generalisability of these findings may be limited as the focus of the study is a specific cohort of students enrolled in blended learning programmes who spend relatively less time on campus. Furthermore, as engagement was operationalised mainly through LMS trace data and predictions were generated from this source, the predictive models represented student behaviour primarily in a digital environment and did not account for academic or social engagement that occurs through face-to-face interactions outside of the LMS. Therefore, casual, informal interactions among peers, face-to-face advising, and participation in extracurricular activities, which have been shown to have an impact on a student's persistence, were not directly observed. Additionally, institutional context and the centralised nature of higher education may affect engagement and persistence in a manner that may differ from other national systems. Finally, due to the focus of this study on academic risk at the course level rather than withdrawal from an institution over time, the results should be interpreted with caution.

Conclusion

The aim of this study was to better understand how predictive modelling could identify behaviour patterns related to student achievement in a blended learning context and provide guidance for developing retention techniques in the future. The study included a number of machine learning models, such as Random Forests, Logistic Regression, XGBoost and MLP that used engagement data taken from a learning management system. The results gave important information on the efficacy and restrictions of these various models.

In terms of educational management, this study has created a scalable and theory-based decision-making framework for institutions. By using data related to LMS usage, class attendance and quiz scores, which have been validated through both global and local model interpretability, predictive models like these can be integrated into support systems by providing students with early warnings about their academic performance, providing academic advisors with the necessary information on how to help students succeed, and giving faculty members access to data to help adjust their instructional methods accordingly. Educational leaders will be able to strategically allocate institutional resources, help to identify and mitigate student risk on a larger scale, and maintain performance and retention standards through the use of such predictive models. However, ethical issues will also need to be considered to ensure that predictive models are adopted by institutions. Future efforts to implement predictive models must address issues related to governance, transparency, fairness, and ethical considerations so that predictive systems enhance – and do not replace – the professional judgement of educators. The alignment of predictive attributes with educational theories increases the likelihood that administrative personnel will find predictive models to be both actionable and grounded in theory.

Although there is substantial potential through the use of machine learning to increase student retention and enhance the responsiveness of institutions, successful deployment will require using both substantial analytic capabilities and the contextual and relational knowledge possessed by educators. This study creates a framework for the use of both analytic and contextual knowledge and supports the movement towards proactive, data-driven leadership in higher education. This study adds to the existing literature supporting the use of learning analytics-based interventions in blended educational environments and it is particularly relevant for the strategic management of education. It emphasises the importance of:

- Continued research on sampling techniques and modelling strategies to provide a better response to class imbalance thus improving the reliability of institutional decision-makers' risk identification efforts.
- Continued investigation into explainable AI tools to increase the transparency of predictive programmes and create confidence amongst stakeholders in the predictive systems that are used to advise and retain students.
- Interdisciplinary teaming among academic staff, data scientists and institutional leaders to create predictive frameworks that are both proven technically and provide educationally sound and contextually appropriate predictive analytical frameworks.
- Application of advanced regularisation techniques such as dropout and L2 regularisation penalties (methods to improve a model's ability to generalise by limiting the influence of less important data and reducing the risk of overfitting to specific patterns) to improve generalisability and ensure consistency across different institutional contexts.

From an educational leadership perspective, effective implementation of predictive analytics requires the combination of robust data-based diagnostics with the contextualised intelligence of educational leaders, advisors, and student affairs professionals. This hybrid strategy is particularly appropriate for developing and implementing proactive retention strategies, identifying student needs on a broad scale and aligning interventions with institutional goals for equity and academic success. The proposed conceptual framework presented in Figure 1 below provides a comprehensive model of the processes required to successfully predict student success in blended model learning environments.

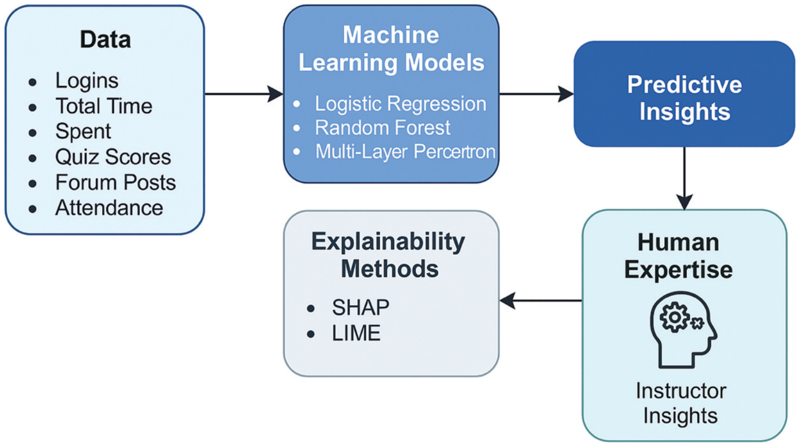

Figure 1. Conceptual framework for predicting student success in blended learning environments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

İrfan Ayhan http://orcid.org/0000-0002-6111-6555

Availability of data and material

The data used in this study were collected from three higher education institutions under formal data use agreements and include sensitive student information. In accordance with institutional regulations and Article 4, Clause 2(g) of the Scientific Research and Publication Ethics Directive of the Turkish Council of Higher Education (YÖK), the raw dataset cannot be shared publicly.

Researchers who wish to access anonymised sample data or summary statistics may contact the corresponding author to initiate a formal request. Access can only be granted following ethical review and written approval from all contributing institutions.

References

Almalawi, A., Soh, B., Li, A., & Samra, H. (2024). Predictive models for educational purposes: A systematic review. *Big Data and Cognitive Computing*, *8*(12), 187. https://doi.org/10.3390/bdcc8120187

Althaqafi, T., Saleem, F., & AL-Ghamdi AL-Malaise, A. (2025). Enhancing student performance prediction: The role of class imbalance handling in machine learning models. *Discover Computing*, *28*(1). https://doi.org/10.1007/s10791-025-09576-4

Anthony, B., Kamaludin, A., Romli, A., Mat Raffei, A. F., Nincarean, D., Eh Phon, A. L., Abdullah, A., & Gan, L. L. M. (2020). Blended learning adoption and implementation in higher education: A theoretical and systematic review. *Technology, Knowledge, and Learning*, *27*(2), 531–578. https://doi.org/10.1007/s10758-020-09477-z

Berens, J., Diem, A., Rumert, L., Schneider, K., & Wolter, S. C. (2025). Crossing individual university boundaries: A comprehensive approach to predicting dropouts in the higher education system. *Higher Education*. https://doi.org/10.1007/s10734-025-01509-w

Carballo-Mendivil, B., Arellano-Gonzalez, A., Rios-Vazquez, N. T., & Lizardi-Duarte, M. D. (2025). Predicting student dropout from day one: XGBoost-based early warning system using pre-enrollment data. *Applied Sciences*, *15*(16), 9202. https://doi.org/10.3390/app15169202

Chang, Y.-H., Chen, F.-C., & Lee, C.-I. (2025). Developing an early warning system with personalized interventions to enhance academic outcomes for at-risk students in Taiwanese higher education. *Education Sciences*, *15*(10), 1321. https://doi.org/10.3390/educsci15101321

De Bruijn-Smolders, M., & Prinsen, F. (2024). Effective student engagement with blended learning: A systematic review. *Heliyon*, *70*(23), e39439. https://doi.org/10.1016/j.heliyon.2024.e39439

Deschacht, N., & Goeman, K. (2015). The effect of blended learning on course persistence and performance of adult learners: A difference-in-differences analysis. *Computers and Education*, *87*, 83–89. https://doi.org/10.1016/j.compedu.2015.03.020

Elibol, S., & Bozkurt, A. (2025). Unfulfilled promises: The complex reasons behind dropout in open online distance education in Türkiye. *American Journal of Distance Education*, 1–16. https://doi.org/10.1080/08923647.2025.2510736

Fan, S., Trimble, A., Kember, D., Muir, T., Douglas, T., Wang, Y., Masters, J., & Mainsbridge, C. (2024). Supporting engagement and retention of online and blended-learning students: A qualitative study from an Australian university. *The Australian Educational Researcher*, *51*(1), 403–421. https://doi.org/10.1007/s13384-022-00605-5

Fredricks, J.A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gandara, D., Anahideh, H., Ison, M. P., & Picchiarini, L. (2024). Inside the black box: Detecting and mitigating algorithmic bias across racialized groups in college student-success prediction. *AERA Open*, *10*. https://doi.org/10.1177/23328584241258741

Goh, T.-T. (2025). Learning management system log analytics: The role of persistence and consistency of engagement behaviour on academic success. *Journal of Computers in Education*, *13*(1), 283–306. https://doi.org/10.1007/s40692-025-00358-x

Guan, X., Feng, X., & Islam, A. (2023). The dilemma and countermeasures of educational data ethics in the age of intelligence. *Humanities and Social Sciences Communications*, *10*(1), 138. https://doi.org/10.1057/s41599-023-01633-x

Guevara-Reyes, R., Ortiz-Garces, I., Andrade, R., Cox-Riquetti, F., & Villegas-Ch, W. (2025). Machine learning models for academic performance prediction: Interpretability and application in educational decision-making. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1632315

Guzmán-Valenzuela, C., Gómez-González, C., Rojas-Murphy Tagle, A., & Lorca-Vyhmeister, A. (2021). Learning analytics in higher education: A preponderance of analytics but very little learning? *International Journal of Educational Technology in Higher Education*, *18*(1). https://doi.org/10.1186/s41239-021-00258-x

Hawthorne, M.J., Zhang, A., & Cooper, A. (2021). Advising undergraduate students: An exploration of how academic advising impacts student success. *Research in Higher Education*, *41*. https://eric.ed.gov/?id=EJ1347870

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-al complementarity. *Journal of Learning Analytics*, *6*(2), 27–52. https://doi.org/10.18608/jla.2019.62.3

Ifenthaler, D., & Yau, I. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research & Development*, *68*(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

İlkdoğan Serbes, D., & Balyer, A. (2025). Dropout tendencies in Turkish higher education. *Problems of Education in the 21st Century*, *83*(4), 545–562. https://doi.org/10.33225/pec/25.83.545

Jonassen, D. (1994). Thinking technology: Toward a constructivist design model. *Educational Technology*, *34*(4), 34–37.

Karademir Coşkun, T., & Alper, A. (2024). Evaluating the evaluators: A comparative study of AI and teacher assessments in higher education. *Digital Education Review*, *45*(45), 124–140. https://doi.org/10.1344/der.2024.45.124-140

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kok, C.L., Ho, C. K., Chen, L., Koh, Y. Y., & Tian, B. (2024). A novel predictive modeling for student attrition utilising machine learning and sustainable big data analytics. *Applied Sciences*, *14*(21), 9633. https://doi.org/10.3390/app14219633

Law, T.-J., Ting, C.-Y., Ng, H., Goh, H.-N., & Quek, A. (2024). Ensemble-SMOTE: Mitigating class imbalance in graduate on time detection. *Journal of Informatics and Web Engineering*, *3*(2), 229–250. https://doi.org/10.33093/jiwe.2024.3.2.17

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education- a literature review. In A. Pena-Ayala (Ed.), *Learning analytics: Fundaments, applications, and trends. Studies in systems, decision and control* (Vol. 94). Springer. https://doi.org/10.1007/978-3-319-52977-6_1

Li, L., & Zeng, D. (2025). Effect of teacher autonomy support on student engagement in physical education classrooms in a blended learning environment: The mediating role of performance expectancy and academic self-efficacy. *BMC Psychology*, *13*(1). https://doi.org/10.1186/s40359-025-02685-1

Liu, Y., Ma, S., & Chen, Y. (2024). The impacts of learning motivation, emotional engagement and psychological capital on academic performance in a blended learning university course. *Frontiers in Psychology*, *15*, 1357936. https://doi.org/10.3389/fpsyg.2024.1357936

Macfadyen, L.P., & Dawson, S. (2012). Numbers are not enough. Why e-Learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology and Society*, *15*(3), 149–163. http://www.jstor.org/stable/jeductechsoci.15.3.149

Millimouno, T.M., Delvaux, T., Kolié, J. M., Kourouma, K., Van Bastelaere, S., Tsunami, C. K., Béavogui, A. H., Garcia, M., Van Damme, W., & Delamou, A. (2022). Evaluation of three blended learning courses to strengthen health professionals' capacity in primary health care, management of sexual and reproductive health services and research methods in Guinea. *Frontiers in Digital Health*, *4*, 911089. https://doi.org/10.3389/fdgth.2022.911089

Mishara, P. (2024). The ethical implications of AI in education: Privacy, bias, and accountability. *Journal of Informatics Education and Research*, *4*(2). https://doi.org/10.52783/jier.v4i2.1827

Nguyen, P.B., & Dao, N. (2024). Exploring learner autonomy in blended learning: A mixed-methods study of English language teaching students in Southern Vietnam. *Journal of Organizational Behavior Managements^)*, 203–235. https://doi.org/10.5281/zenodo.12543009

Ninaus, M., & Sailer, M. (2022). Closing the loop - the human role in artificial intelligence for education. *Frontiers in Psychology*, *13*, 956798. https://doi.org/10.3389/fpsyg.2022.956798

OECD. (2025). *Education at a glance, 2025: Türkiye*. Retrieved from OECD: https://www.oecd.org/en/publications/education-at-a-glance-2025_1a3543e2-en/turkiye_9648f416-en.html

Opazo, D., Moreno, S., Alvarez-Miranda, E., & Pereira, J. (2021). Analysis of first-year university student dropout through machine learning models: A comparison between universities. *Mathematics*, *9*(20), 2599. https://doi.org/10.3390/math9202599

Paolucci, C., Vancini, S., Bex, R. T., II, Cavanaugh, C., Salama, C., & de Araujo, Z. (2024). A review of learning analytics opportunities and challenges for K-12 education. *Heliyon*, *10*(4), e25767. https://doi.org/10.1016/j.heliyon.2024.e25767

Powers, D. (2011). Evaluation: From precision, recall and F-Measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, *2*(1), 37–63.

Pratama, F.R., Santoso, H. B., Junus, K., Michael, J., Mannix, I. A., & Athaya, H. (2025). Constructivism in online and hybrid learning before and after COVID-19: A systematic literature review. *Jurnal Eduscience (JES)*, *12*(4), 1–22. https://doi.org/10.36987/jes.v12i4.7314

Rahmani, A.M., Groot, W., & Rahmani, H. (2024). Dropout in online higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00450-9

Rebelo Marcolino, M., Porto, T. R., Primo, T. T., Targino, R., Ramos, V., Queiroga, E. M., Cechinel, C., & Cechinel, C. (2025). Student dropout prediction through machine learning optimization: Insights from Moodle log data. *Scientific Reports*, *15*(1), 9840. https://doi.org/10.1038/s41598-025-93918-1

Santos, R.M., & Henriques, R. (2023). Accurate, timely, and portable: Course-agnostic early prediction of student performance from LMS logs. *Computers and Education: Artificial Intelligence*, *5*, 100175. https://doi.org/10.1016/j.caeai.2023.100175

Shavelson, R.I., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz- Primo, M. A., Yin, Y., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, *21*(4), 295–314. https://doi.org/10.1080/08957340802347647

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

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

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, *19*(1), 12. https://doi.org/10.1186/s41239-021-00313-7

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press.

Tiwari, M., & Jain, N. (2024). Student performance prediction using machine learning algorithms. *ShodhKosh: Journal of Visual and Performing Arts*, *5*(6), 1112–1122. https://doi.org/10.29121/shodhkosh.v5.i6.2024.4552

Tsai, Y.-S., Poquet, O., Casevic, D., Dawson, S., & Pardo, A. (2019). Complexity leadership in learning analytics: Drivers, challenges, and opportunities. *British Journal of Educational Technology*, *50*(6), 2839–2854. https://doi.org/10.1111/bjet.12846

Usta Ergün, S.B. (2025). A hybrid approach to analyzing factors influencing international student retention in Turkish higher education: Integrating fuzzy DEMATEL, machine learning, and statistical methods. *Journal of Fuzzy Extension and Applications*, *6*(2), 233–259. https://doi.org/10.22105/jfea.2024.484077.1668

Villar, A., & de Andrade, C. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Discover Artificial Intelligence*, *4*(1). https://doi.org/10.1007/s44163-023-00079-z

Vygotsky, V. L. (1978). *Mind in society: Development of higher psychological processes*. https://doi.org/10.2307/j.ctvjf9vz4

Waheed, H., Hassan, S., Aljohani, N., Hardman, L., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VUE big data using deep learning models. *Computers in Human Behavior*, *104*, 106189. https://doi.org/10.1016/j.chb.2019.106189

Whitelock-Wainwright, A., Tsai, Y.-S., Lyons, K., Khalif, S., Bryant, M., Ryan, K., & Casevic, D. (2020). Disciplinary differences in blended learning design: A network analytic study. LAK '20: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 579–588). Frankfurt, Germany: Association for Computing Machinery (ACM). https://doi.org/10.1145/3375462

YÖKAK. (2025). *Durum Raporları - Yükseköğretim Kalite Kurulu Başkanlığı*. Retrieved from Yükseköğretim Kalite Kurulu: https://www.yokak.gov.tr/documents/StatusReports/2024%20Yılı%20Yükseköğretim%20Değerlendirme%20ve%20Kalite%20Güvencesi%20Durum%20Raporu%20.pdf

Zawacki-Richter, O., Marin, V., Bond, M., & Gouvemeur, F. (2019). Systematic review of research on artificial intelligence applications in higher education - where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1). https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B.J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *47*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2