OPUS
Journal of Society Research

Type: Research Article
Received Date: May 11, 2026
Accepted Date: August 11, 2026
Published: August 12, 2026
10.26466/opusjsr.1948539

Early identification of at-risk students in online learning environments: A learning analytics approach using machine learning models

Çevrimiçi öğrenme ortamlarında risk altındaki öğrencilerin erken belirlenmesi: Makine öğrenimi modelleri kullanan bir öğrenme analitiği yaklaşımı

Verda Gizem Oğul[1] **Yavuz Selim Balcıoğlu[2]**

[1] *Lecturer, Marmara University, İstanbul, Türkiye*

[2] *Assoc. Prof. Dr., Doğuş University, Department of Management Information Systems, İstanbul, Türkiye*

Corresponding Author:
Verda Gizem Oğul

Citation:
Oğul, V. G. & Balcıoğlu, Y. S. (2026). Early identification of at-risk students in online learning environments: A learning analytics approach using machine learning models. *OPUS– Journal of Society Research, 23,* e1948539.
https://doi.org/10.26466/opusjsr.1948539

Review Note:
Evaluated by Double-Blind Peer Review

Ethics Reporting:
To report potential ethical concerns, contact:
editorialoffice@opusjournal.net

Similarity screening was conducted via intihal.net.

Abstract

Online learning environments generate temporally ordered traces that can support early academic-risk screening, but credible claims require an explicit prediction point and strict control of post-outcome information. This study evaluates whether information available by day 28 of an Open University module presentation can identify students who subsequently fail or withdraw. Raw tables from the Open University Learning Analytics Dataset (OULAD) were reconstructed at student–module–presentation level after timestamped Virtual Learning Environment (VLE) and assessment records had been truncated at the prediction point. The day-28 risk set comprised 27,522 enrolments from 24,832 unique students; 44.1% subsequently failed or withdrew. Seven classifiers were compared using student-grouped five-fold validation, randomized hyperparameter optimization, and a student-disjoint held-out test set. Gradient Boosting achieved the highest grouped-validation ROC-AUC and yielded test ROC-AUC=0.790 (95% CI: 0.777–0.801), PR-AUC=0.770 (95% CI: 0.753–0.786), accuracy=0.724, precision=0.728, recall=0.597, and F1=0.656. Performance increased from day 14 (ROC-AUC=0.747) to day 56 (0.835), demonstrating an earliness–performance trade-off rather than uniformly high early accuracy. SHAP analysis identified completion of assessments due by the cutoff, assessment performance, prior education, and VLE engagement as influential predictors. A score-excluded specification retained meaningful discrimination (ROC-AUC=0.762). Three-class sensitivity analysis showed that failure and withdrawal were considerably harder to distinguish than success. The findings support cautious use of day-28 risk scores for prioritizing supportive outreach, while emphasizing calibration, subgroup monitoring, human oversight, and local validation.

Keywords: *Online learning, learning analytics, early warning systems, machine learning, OULAD, explainable artificial intelligence*

Öz

Çevrim içi öğrenme ortamlarında üretilen zamansal izler, akademik riskin erken belirlenmesini destekleyebilir; ancak güvenilir bir erken tahmin iddiası, açık bir tahmin noktası ve sonuç sonrası bilgilerin sıkı biçimde dışlanmasını gerektirir. Bu çalışma, Açık Üniversite Öğrenme Analitiği Veri Setinde (OULAD) bir modül sunumunun 28. gününe kadar mevcut olan bilgilerin daha sonra başarısız olan veya kaydını sildiren öğrencileri belirleyip belirleyemediğini incelemektedir. Ham OULAD tabloları, VLE ve değerlendirme kayıtları tahmin noktasında kesildikten sonra öğrenci–modül–sunum düzeyinde yeniden oluşturulmuştur. 28. gün risk kümesi 24.832 farklı öğrenciye ait 27.522 kayıttan oluşmuş; bunların %44,1'i daha sonra başarısız olmuş veya kaydını sildirmiştir. Yedi sınıflandırıcı, öğrenci gruplu beş katlı doğrulama, rastgele hiperparametre optimizasyonu ve öğrenciler bakımından bağımsız bir test kümesiyle karşılaştırılmıştır. Gradyan Artırma modeli test kümesinde ROC-AUC=0,790 (%95 GA: 0,777–0,801), PR-AUC=0,770 (%95 GA: 0,753–0,786), doğruluk=0,724, kesinlik=0,728, duyarlılık=0,597 ve F1=0,656 sonuçlarını vermiştir. Performans 14. günden (ROC-AUC=0,747) 56. güne (0,835) yükselmiş ve erkenlik ile performans arasındaki ödünleşimi göstermiştir. SHAP analizi, vadesi gelmiş değerlendirmelerin tamamlanma oranı, değerlendirme performansı, önceki eğitim düzeyi ve VLE katılımını öne çıkarmıştır. Değerlendirme puanlarının dışlandığı model de anlamlı ayırt edicilik göstermiştir (ROC-AUC=0,762). Üç sınıflı duyarlılık analizi, başarısızlık ile kayıttan çekilmenin başarıya kıyasla daha güç ayrıldığını göstermektedir. Bulgular, 28. gün risk skorlarının destekleyici müdahaleleri önceliklendirmek amacıyla temkinli kullanımını desteklemekte; kalibrasyon, alt grup izleme, insan denetimi ve yerel doğrulama gereğini vurgulamaktadır.

Anahtar Kelimeler: *Çevrim içi öğrenme, öğrenme analitikleri, erken uyarı sistemleri, makine öğrenmesi, OULAD, açıklanabilir yapay zekâ*

Introduction

In the last 20 years, the explosion of online learning has transformed the way universities and other educational providers deliver their programs. Many colleges now use online delivery as their primary means of instruction (rather than simply offering it as a secondary option), which was further accelerated by the COVID-19 pandemic (Basilaia & Kvavadze, 2020; Hodges et al., 2020). One of the challenges created by this rapid transition is identifying at-risk students (those who are likely to either fail or withdraw from their program) in order to provide them with timely assistance. The ability to detect students who may ultimately struggle will serve as an essential first step in providing appropriate support in a timely manner and helping to create a more equitable learning environment for all students (Huptchy et al., 2018; Ferguson, 2013).

Adult and continuing education is a sector that is intrinsically tied to the concept of lifelong learning, with online platforms now playing a critical role within this sector (Dabbagh & Fake, 2017). Online courses, such as MOOCs, hybrid programs, and workplace training courses are increasingly attracting students from diverse socioeconomic backgrounds; therefore, these students also have diverse motivations for enrolling in courses, various levels of previous education, and various levels of institutional support. Therefore, the rate of dropouts and failures from courses offered in this manner is typically much greater than those offered in a traditional classroom setting. It is necessary to identify at-risk students early enough to provide appropriate support. Using data-based approaches allows us to identify students who may experience difficulty prior to their difficulties becoming irreversible; Jordan (2014) also states this as well as Onah et al. (2014).

The urgency of supporting at-risk learners has grown considerably in the years following the widespread shift to digital education. Research conducted in the post-pandemic period has underscored that the challenges of student retention and academic performance are not temporary disruptions but rather structural features of online learning environments that require systematic, data-driven responses (Viberg et al., 2024). Recent studies have demonstrated that students from lower socioeconomic backgrounds and first-generation learners are disproportionately affected by disengagement in online settings, reinforcing the need for equitable, proactive identification mechanisms (Herodotou et al., 2023). Furthermore, the integration of artificial intelligence into educational platforms has opened new possibilities for continuous, real-time monitoring of student behavior, reducing the delay between the emergence of risk indicators and institutional response (Aldowah et al., 2019; Bravo et al., 2025).

To assist with this issue, the field of learning analytics has emerged as a viable means of research for addressing these issues. As a general definition, learning analytics is the process of measuring (collecting), analyzing, and reporting data about students and their learning environments in order to improve those students' outcomes (Siemens & Baker, 2012). Essentially, learning analytics involves tracking and understanding how students interact with their online courses through various digital traces left behind by the students, such as when a student has logged into an online course, submitting an assignment, or accessing course materials. In addition to this analysis of digital traces, machine learning has provided a new level of sophistication and accuracy to predict students' success and ease of use in educational practice. Therefore, institutions can develop comprehensive early warning systems that can serve large numbers of learners (Arnold & Pistilli, 2012; Essa & Ayad, 2012).

Risk prediction with OULAD is not itself a new research problem. Prior studies have examined weekly prediction, course-percentage cutoffs, ensemble learning, deep learning, and explainable models. The remaining methodological problem is comparability and credibility: published performance estimates depend strongly on the prediction time, target definition, feature-availability rule, enrolment structure, and validation design. Models built from whole-course aggregates or withdrawal-related fields may describe outcomes well but cannot support a genuine early-warning claim. Accordingly, this study does not claim to introduce OULAD risk prediction; it contributes a transparent day-based reconstruction, student-grouped validation, leakage auditing, calibration and uncertainty analysis, SHAP explanations, and explicit sensitivity tests.

The study therefore asks whether an operationally useful risk signal is available at day 28, how performance changes at days 14, 42, and 56, how seven commonly used classifiers compare under identical features and splits, and which predictors drive the selected model. The contribution lies in the combined evaluation protocol rather than in a claim that OULAD or machine-learning risk prediction has not previously been studied.

Responsible use requires more than high discrimination. The revised analysis therefore reports probability calibration, student-cluster confidence intervals, learning curves, course-presentation holdout performance, score- and demographic-exclusion specifications, subgroup diagnostics, and SHAP explanations. These analyses are diagnostic rather than proof of fairness or causal effects; risk scores are intended to support human-led outreach, not autonomous adverse decisions.

2. Literature Review

2.1 Learning Analytics in Online and Distance Education

The study of learning analytics has expanded rapidly within academia and as a practical means of utilizing educational technology. To analyze data generated by the student, computational methods are used with the intention of providing educational institutions with data that can demonstrate the relationship between different engagement processes (i.e., how many students completed a specific assignment). This kind of study offers insight into factors (behavioral) which can lead to student success.

For example, through the use of engaging engagement data collected from LMS, researchers have been able to create a detailed picture of how individual learners engage with their respective modules (Gasevic et al., 2016). Researchers have established that data on engagement provided by LMS can be reliable predictors of final grades and course completion across platforms such as Moodle, Blackboard, and Coursera (Macfadyen & Dawson, 2010; Romero et al., 2013; Romero & Ventura, 2020).

When analyzing publicly funded and continuing education, researchers are faced with different challenges than when analyzing traditional undergraduate students. Adult learners have non-traditional pathways and may be working and taking care of families at the same time as they are enrolled in online programs (Merriam & Bierema, 2014; Papamitsiou & Economides, 2014). Consequently, the definition of engagement that applies to traditional undergraduate learners may not be accurate in the world of continuing education. Researchers in adult education have noted that continuing education students possess diverse motivations for continuing their education and have many circumstances that need to be considered when using learning analytics frameworks (Drachsler & Greller, 2016; Rienties & Toetenel, 2016).

2.2 Predictive Modeling and Machine Learning for At-Risk Student Identification

Numerous studies show that for over a decade researcher have used machine learning to predict which students are at risk of failing or dropping out of school. Many of these researchers have used supervised classification algorithms (i.e., decision trees, artificial neural networks, support vector machines, etc.) to build predictive models using data from schools. In many instances, studies comparing different supervised classification techniques have shown that ensemble techniques (Random Forest and Gradient Boosting techniques) produce more accurate predictions than less sophisticated algorithms (e.g., Logistic Regression); however, the relative performance of these two techniques will depend on both the structure of the educational dataset being used as well as the characteristics included in the model (Mduma et al., 2019).

OULAD is a widely used benchmark containing linked demographic, registration, assessment, module, and daily VLE interaction tables for 32,593 enrolments in 22 presentations of seven anonymized modules (Kuzilek et al., 2017). This breadth enables direct comparison, but reported performance is highly design-dependent. Hlosta et al. (2017) predicted near-term assessment non-submission at daily cutoffs; Jha et al. (2019) compared ensemble, deep-learning, and regression models for dropout and final result; Adnan et al. (2021) evaluated risk at several percentages of course length; Al-Zawqari et al. (2022) emphasized flexible feature selection; Zhang and Ahn (2023) used semester-aware weekly prediction and SHAP; and Ujkani et al. (2024) combined multiple classifiers with SHAP. These studies establish that OULAD prediction already exists and also demonstrate why accuracy values cannot be compared without aligning the target, prediction horizon, feature availability, and validation unit.

2.3 Lifelong Learning, Public Education and Digital Inclusion

The concept of lifelong learning (LLL) encompasses all three types of learning: formal (i.e., classroom-based), nonformal (e.g., community college), and informal (e.g., field experience), across all phases of life. Over the past two decades, LLL has been rising in importance in educational policy debates (European Commission, 2020; UNESCO, 2016). Global changes in employment markets, rapid advances in technology, and a growing recognition of the limitations of traditional front-loaded education systems have led governments and international organizations to promote learning systems that support continual development of skills and civic participation throughout adult life. Open and distance educational institutions have emerged, in this context (as recognized in Zawacki-Richter et al. [2019]), as key providers of accessible, convenient, and affordable lifelong learning opportunities to groups that have historically not had equal access to postsecondary education.

Today, most LLL is delivered through digital platforms and digital learning environments. Online learning environments have many advantages, including reach, accessibility, the ability to offer flexible hours, and lower costs. This allows remote and low-income learners who would not otherwise have had access to the educational materials they need for career advancement or citizenship (Zawacki-Richter & Naidu, 2016). However, the accelerating transition to digital delivery of education has created new forms of inequity in education because an individual's participation in an online learning environment is impacted by access to technology (computers and the internet) and an individual's ability to learn the requisite skills to navigate a digital learning environment (Tsai et al., 2020).Initially, the concept of the digital divide was focused on technology infrastructure access. In recent years, however, more attention has been paid to the issues of ability, motivation to use technology, and patterns of technology usage and how those factors are related to educational outcomes.

There have been increasing discussions in public education systems about how the use of learning analytics can help to create increased equity and inclusion. For instance, at-risk student early warning systems can help educational institutions direct the provision of support resources to the students who need them the most, thereby resulting in fairer education outcomes (Law & Liang, 2020). Many educational institutions that have developed programs based on learning analytics report significant positive impacts on student persistence, you will also find that completion rates for first-generation college students and non-traditional learners (Arnold & Pistilli, 2012; Sclater et al., 2016). The effectiveness of the early warning systems for detecting at-risk students will be influenced by the quality of the predictive models used and the ability of the institution to provide timely and targeted support based on the findings generated through analytics.

There's been new academic research to suggest that learning analytics projects should be viewed within a larger ethical and policy framework that appropriately addresses the considerations of data governance, learners giving consent, and institutional responsibility (Tsai et al., 2020; Williamson et al., 2020).The Public education services entail sensitivity level of caregivers and students due to several reasons, such as; students enrolled in Adult Education and Continuing Education programs are likely to have social vulnerabilities; students may have concerns regarding the types of data collected. Thus; making use of predictive analytics in Public Education environments requires responsible use of predictive analytics; this requires more than technical and quality aspects. It requires a strong commitment to transparency regarding how systems operate; it requires the involvement of learners during the design phase of systems; and it requires ongoing evaluations of the broad impact of decisions made by algorithms.

2.4 Feature Selection and Interpretability of Educational Predictive Models

Among the most difficult aspects of creating learning analytics models, is selecting predictor variables that are useful to create future predictions and are easily interpreted by education professionals. The existence of large volumes of data in an LMS creates a risk of overfitting (i.e. adding too many predictor variables) or multicollinearity (i.e. two or more predictor variables are highly correlated) if there is no established variable selection method (Baker & Inventado, 2014). This is a challenge faced by researchers who have attempted to address historically in many different ways; including, filter methods - which use statistical associations; wrapped methods - which evaluate different combinations of predictor variables based on predictive performance; embedded methods (e.g., calculating feature importance based on tree-based ensembles) (Kotsiantis, 2012).

Interpretability of predictive models is especially important in education because those responsible for making decisions, require trust and understanding when implementing decisions based on the recommendations of algorithms (Rudin, 2019). Making accurate predictions using the model but not being able to interpret the reason for the prediction is problematic from a practical perspective; even if the predictive ability of the model is better than those of simpler models based on various technical criteria. As a result, researchers in learning analytics are devoting greater attention to Explanatory Artificial Intelligence (XAI) methods (e.g., SHAP and LIME), which provide general and individual level explanations of the predictions made by models (Lundberg & Lee, 2017).

Research indicates that the use of feature importance analysis produces results that are meaningful educationally. Feature importance analysis consistently identified assessment-related variables (e.g., average assessment score; frequency of submission of assessments; variance of assessment scores) and engagement timing measures to be the highest predictive powers for students in online learning data (Hlosta et al. 2017; Hussain et al. 2018). These findings support the theoretical models of self-regulated learning which suggest that academic success is most influenced by the processes of metacognitive monitoring and strategic effort (Zimmerman, 2000). The knowledge of which features are the most predictive assists the education sector in targeting their monitoring and intervention efforts at the behavioral signals most useful in identifying students 'at-risk' (Kuzilek et al., 2017).

Research Questions:

RQ1: Which learning analytics indicators are most strongly associated with students' academic risk in online learning environments?

RQ2: To what extent can machine learning models accurately predict at-risk students using learning analytics data?

RQ3: Which machine learning model provides the best predictive performance for identifying at-risk students in online learning environments?

RQ4: Which behavioral and assessment-related features contribute most strongly to the prediction of student academic risk?

2.5 Direct Comparison with OULAD Studies

Table 1 compares studies that are particularly relevant to the revised design. The comparison shows that the present contribution is methodological: a fixed day-28 prediction point, pre-aggregation temporal truncation, student-grouped validation, uncertainty and calibration reporting, score-availability sensitivity, subgroup diagnostics, and SHAP-based interpretation are combined in one reproducible workflow.

Table 1. Direct comparison of the present study with selected OULAD prediction studies

Study	Objective/method	Main finding	Main limitation for comparison
Hlosta et al. (2017)	Daily prediction of assessment non-submission; LR, SVM, RF, NB, XGB	PR-AUC varied markedly by day; timing affected utility	Different target; course-specific daily cutoffs
Jha et al. (2019)	Dropout and result prediction; GBM, DRF, DL, GLM	AUC up to 0.91 for dropout and 0.93 for result	Whole-record feature construction; not a fixed early cutoff
Adnan et al. (2021)	Risk prediction at 20–100% of course length using multiple ML models	Performance increased as more course data became available	Percentage-based horizons; comparison depends on course length
Al-Zawqari et al. (2022)	RF and ANN with flexible feature selection	86–88% accuracy for failure; up to 93% for dropout	Class-specific tasks; timing and validation differ
Zhang & Ahn (2023)	Semester-aware weekly prediction; 11 single/integrated algorithms; SHAP	F1 generally improved from weeks 8 to 32; voting reached 84% end-semester accuracy	Course/semester-specific models; later horizons
Ujkani et al. (2024)	ML/DL at-risk classification with SHAP	Best custom NN accuracy 93%; SHAP highlighted engagement and registration	Prediction horizon and grouping differ from the present design
Present study	Days 14/28/42/56; seven ML models; student grouping; SHAP; calibration and sensitivity analyses	Day-28 test ROC-AUC 0.790; PR-AUC 0.770; F1 0.656	Single institution; score-release timing unavailable; no intervention evaluation

3. Methodology

3.1 Data Source, Scope, and Unit of Analysis

The analysis used the official Open University Learning Analytics Dataset (OULAD), obtained from the Open University repository and documented by Kuzilek et al. (2017). OULAD contains anonymized data from 2013–2014 for seven modules and 22 module presentations. The raw files comprise 32,593 enrolment rows, 28,785 unique students, 173,912 student-assessment records, and 10,655,280 VLE interaction rows. The unit of analysis was a student–module–presentation enrolment. The anonymized module codes do not identify a programme, undergraduate/postgraduate status, or formal academic level; therefore, these attributes cannot be inferred and are not claimed in this study. The variable highest_education describes the learner's prior qualification, not the level of the module.

OULAD was selected for methodological comparability rather than as a proxy for contemporary online-learning populations. Its publicly documented relational structure links registration, demographic, assessment, course, and daily VLE-interaction records, allowing prediction-time features to be reconstructed transparently and compared directly with prior OULAD studies. The 2013-2014 observation period is therefore appropriate for evaluating the reproducibility and leakage control of the analytical protocol, but it cannot establish the current institutional validity of the fitted model.

The primary prediction point was day 28 relative to the official module start. Sensitivity analyses used days 14, 42, and 56. The cutoff was an information-availability boundary: timestamped records satisfying date ≤ cutoff were retained before aggregation. Because OULAD module resources can open before the official start, the day-28 dataset contained observable VLE records from day −25 through day 28 (54 calendar indices). Thus, "early" refers to prediction by day 28, not to a claim that exactly 28 calendar days of history were used.

3.2 Table Integration and Enrolment Structure

studentInfo and studentRegistration were merged one-to-one on id_student, code_module, and code_presentation; course metadata were then merged many-to-one by module and presentation. VLE rows were first truncated at the cutoff, aggregated to daily clicks, and then summarized at enrolment level. Assessment submissions were likewise truncated by date_submitted before being joined to assessment metadata and aggregated. Only one-row-per-enrolment feature blocks were merged into the modelling table. No duplicated enrolment keys remained in studentInfo or studentRegistration.

A total of 3,538 students had more than one enrolment. Consequently, all enrolments belonging to the same id_student were assigned to the same train, validation, or test fold. This prevents the model from learning a student's history in one split and evaluating the same student in another.

3.3 Prediction Cohort, Outcome, and Leakage Control

At each cutoff, the operational risk set contained students who had registered and had not already unregistered by that prediction point. The date_unregistration field was used only to define eligibility at the prediction time and was never supplied to a model. Direct or derived unregistration predictors, whole-course activity totals, post-cutoff submissions, and final-result fields were excluded. At day 28, 5,055 already-withdrawn enrolments and 16 not-yet-registered enrolments were excluded, leaving 27,522 eligible enrolments representing 24,832 students.

The primary binary outcome coded Fail and Withdrawn as at risk and Pass and Distinction as not at risk. This grouping reflects the operational decision of whether supportive outreach may be warranted; it does not imply that failure and withdrawal have identical pedagogical causes. A three-class sensitivity analysis separately modelled Success, Fail, and Withdrawn.

3.4 Predictors and Preprocessing

Predictors comprised registration and course information; demographic and prior-study characteristics; overall, recent, and activity-type VLE engagement; and assessment-submission and score summaries observable by the cutoff. The revised completion rate was defined as the number of distinct assessments due and submitted by the cutoff divided by the number due by that cutoff, producing a bounded 0–1 measure. The redundant active_day_rate variable was removed. Numeric variables were median-imputed with missingness indicators and standardized; categorical variables were mode-imputed and one-hot encoded. All preprocessing was estimated inside each training fold.

OULAD records assessment submission dates but not the dates on which scores became institutionally visible. The primary model assumes that scores attached to submissions by the cutoff were available. A conservative score-excluded analysis therefore retained submission behavior but removed all score-derived variables.

3.5 Algorithms, Validation, and Hyperparameter Optimization

Seven algorithms were selected to cover interpretable linear, probabilistic, distance-based, margin-based, single-tree, bagged-tree, and boosted-tree approaches: Logistic Regression, Gaussian Naïve Bayes, k-Nearest Neighbours (k-NN), radial-basis-function Support Vector Machine (SVM), Decision Tree, Random Forest, and Gradient Boosting. Logistic Regression and Naïve Bayes provide transparent baselines; k-NN evaluates local similarity; SVM represents a non-linear margin classifier; Decision Tree provides a simple rule-based comparator; Random Forest evaluates bagging; and Gradient Boosting evaluates sequential ensemble learning. This expanded comparison directly addresses why commonly used alternatives were or were not retained. Table 2 summarizes the model-specific dimensions included in randomized hyperparameter optimization.

One-fold of a five-fold StratifiedGroupKFold formed the held-out test set (5,505 enrolments); the remaining observations were used for development. Five-fold student-grouped validation within the development set supported model comparison and randomized hyperparameter search, using ROC-AUC as the selection criterion. The test set was not used for model or parameter selection. All seven models were tuned. The selected Gradient Boosting configuration used 300 estimators, learning_rate=0.08, max_depth=2, min_samples_leaf=20, min_samples_split=5, subsample=0.70, and random_state=42.

Table 2. Algorithms and randomized hyperparameter-search dimensions.

Model	Randomized-search dimensions
Logistic Regression	C, class weight, solver; L2 penalty
Decision Tree	criterion, depth, minimum split/leaf, class weight
Random Forest	300–800 trees, depth, minimum split/leaf, feature sampling, class weight
Gradient Boosting	100–500 estimators, learning rate, depth, minimum split/leaf, subsample, feature sampling
SVM (RBF)	C, gamma, class weight
Naïve Bayes	variance smoothing
k-NN	neighbours, weights, Minkowski p, leaf size

3.6 Evaluation, Explainability, and Robustness

Accuracy, precision, recall, F1, ROC-AUC, average precision (PR-AUC), and Brier score were reported at a prespecified 0.50 classification threshold. Student-cluster bootstrap resampling (1,000 iterations) provided 95% confidence intervals for the selected model. Calibration was assessed using ten quantile bins. Learning curves compared training and grouped-validation ROC-AUC. Additional analyses comprised course-presentation-grouped holdout validation, demographic- and score-exclusion specifications, subgroup diagnostics, and the three-class outcome model.

Global and directional explanations were generated using SHapley Additive exPlanations (SHAP) for the selected Gradient Boosting model. SHAP values describe associations with model output conditional on the fitted model; they do not establish causal effects. All analyses used Python 3.13.7, scikit-learn 1.9.0, pandas 3.0.3, NumPy 2.2.6, SHAP 0.52.0, and random_state=42

4. Results

4.1 Cohort and Descriptive Statistics

At day 28, the analytical cohort comprised 27,522 enrolments from 24,832 unique students. Of these, 15,379 (55.9%) subsequently passed or achieved distinction and 12,143 (44.1%) failed or withdrew. The raw OULAD records included seven modules and 22 module presentations. Table 3 provides dispersion, central tendency, range, and missingness for the main analytical variables; complete numerical and categorical descriptives for every dependent and independent variable are provided in Supplementary Tables S1-S2. Engagement was highly heterogeneous: total clicks averaged 356.95 (SD=430.90; median=225; range=0-7,697), while active days averaged 14.66 (SD=10.36; median=13; range=0-54). The median learner had last interacted two days before the cutoff. Assessment completion was concentrated at 1.00 but was not applicable for 18.13% of enrolments because no assessment was due by day 28. The weighted mean assessment score was 72.96 (SD=21.57; median=77.47), with 25.90% missing because no observable scored submission was available by the cutoff. Appendix Table A1 provides ten de-identified representative analytical records.

Table 3. Descriptive statistics for the day-28 analytical cohort

Variable	N	Missing (%)	Mean	SD	Median	Min	Max
Active days	27,522	0.00	14.66	10.36	13.00	0.00	54.00
Days since last VLE activity	26,617	3.29	4.00	6.12	2.00	0.00	52.00
Assessment completion rate	22,533	18.13	0.89	0.31	1.00	0.00	1.00
Weighted mean assessment score	20,395	25.90	72.96	21.57	77.47	0.00	100.00
Late submissions	27,522	0.00	0.19	0.39	0.00	0.00	2.00
Prior attempts	27,522	0.00	0.16	0.47	0.00	0.00	6.00
Studied credits	27,522	0.00	76.32	38.24	60.00	30.00	630.00

4.2 Model Comparison and Held-Out Performance

Table 4 reports held-out performance for all seven classifiers. Gradient Boosting had the highest tuned grouped-validation ROC-AUC (0.7919) and was therefore selected without reference to test performance. On the held-out test set it achieved ROC-AUC=0.7895 and PR-AUC=0.7704. Logistic Regression, Random Forest, and SVM produced similar but slightly lower discrimination, indicating that Gradient Boosting's advantage was modest rather than categorical. Naive Bayes had high precision but very low recall at the 0.50 threshold. Figure 1 compares the held-out ROC curves, whereas Figure 2 shows the precision-recall curves and the stronger separation among the leading models relative to the prevalence baseline.

4.3 Uncertainty, Classification Errors, and Calibration

For Gradient Boosting, accuracy was 0.724 (95% CI: 0.711–0.735), precision 0.728 (0.708–0.748), recall 0.597 (0.578–0.617), F1 0.656 (0.639–0.671), ROC-AUC 0.790 (0.777–0.801), PR-AUC 0.770 (0.753–0.786), and Brier score 0.183 (0.178–0.188). The model correctly classified 1,451 of 2,429 at-risk enrolments and missed 978; it generated 542 false-positive alerts among 3,076 not-at-risk enrolments. Therefore, the model is useful for ranking and triage but does not identify "almost all" at-risk students at the default threshold. Table 5 reports the corresponding confusion matrix. Figure 3 shows that observed risk proportions remained close to mean predicted probabilities across deciles, although this retrospective calibration does not replace prospective local validation.

Table 4. Held-out test-set performance of the seven classifiers at day 28

Model	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
Gradient Boosting	0.724	0.728	0.597	0.656	0.790	0.770
Logistic Regression	0.715	0.682	0.666	0.674	0.784	0.764
Random Forest	0.721	0.737	0.571	0.644	0.783	0.762
SVM (RBF)	0.716	0.703	0.617	0.657	0.782	0.748
k-NN	0.709	0.750	0.511	0.608	0.772	0.752
Decision Tree	0.701	0.705	0.552	0.620	0.748	0.719
Naive Bayes	0.641	0.805	0.245	0.376	0.714	0.678

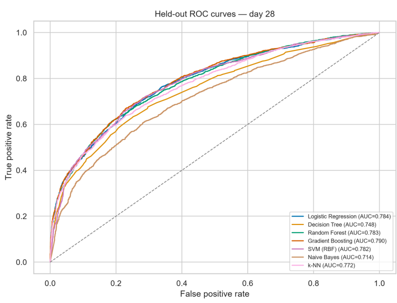

Figure 1. Held-out ROC curves at day 28

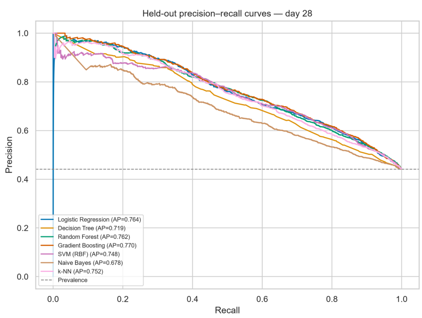

Figure 2. Held-out precision–recall curves at day 28

Table 5. Confusion matrix for the selected Gradient Boosting model at the 0.50 threshold.

	Predicted not at risk	Predicted at risk
Actual not at risk	2,534	542
Actual at risk	978	1,451

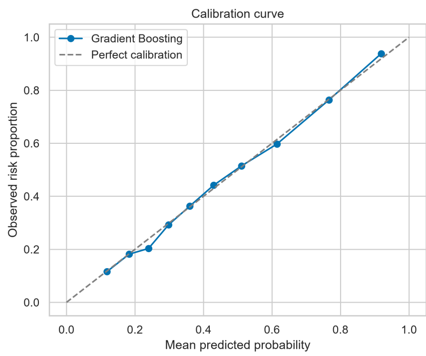

Figure 3. Probability calibration of the selected day-28 model.

4.4 Earliness–Performance Trade-Off

Grouped-validation discrimination improved as later information became available. ROC-AUC increased from 0.747 at day 14 to 0.792 at day 28, 0.812 at day 42, and 0.835 at day 56. F1 increased from 0.628 to 0.697 across the same prediction points. Because students who had already withdrawn were removed from each operational risk set, cohort size and risk prevalence also changed over time; the curves therefore describe performance for the still-enrolled population at each decision point. Table 6 reports the estimates numerically, and Figure 4 visualizes the earliness-performance trade-off.

Table 6. Grouped-validation performance across operational prediction cutoffs.

Cutoff	Enrolments	ROC-AUC	PR-AUC	F1
14	28,061	0.747 ± 0.006	0.713 ± 0.008	0.628 ± 0.005
28	27,522	0.792 ± 0.008	0.772 ± 0.010	0.653 ± 0.015
42	27,012	0.812 ± 0.004	0.791 ± 0.003	0.673 ± 0.005
56	26,513	0.835 ± 0.006	0.816 ± 0.006	0.697 ± 0.009

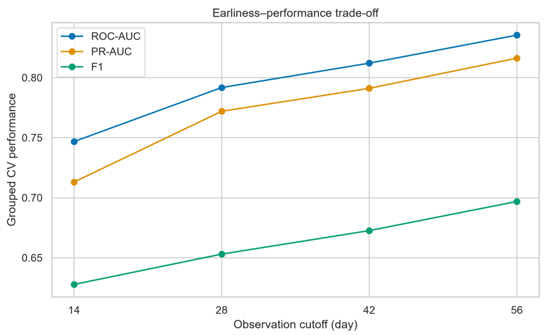

Figure 4. Earliness–performance trade-off across operational cutoffs.

4.5 Overfitting and Course-Presentation Robustness

As shown in Figure 5, the learning curve showed a large training–validation gap with 3,522 observations (0.083) that narrowed to 0.020 at the full development sample. Validation ROC-AUC plateaued near 0.792, suggesting limited residual overfitting and little evidence of underfitting. When entire course-presentation groups were held out, mean ROC-AUC was 0.777±0.028 and PR-AUC was 0.754±0.035. The modest decline relative to ordinary student-grouped validation indicates partial, but not complete, transfer across offerings.

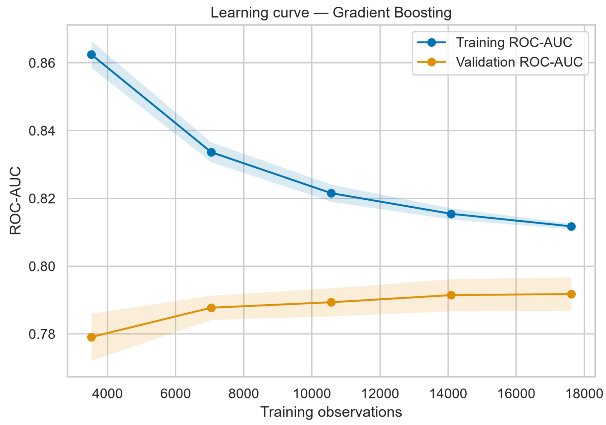

Figure 5. Learning-curve assessment of overfitting and data sufficiency.

4.6 Explainable Artificial Intelligence Results

Figures 6 and 7 present the global and directional SHAP results. Figure 6 ranks predictors by mean absolute SHAP magnitude, identifying assessment completion, weighted assessment score, prior education, active days, module context, click activity, VLE recency, and late submissions as the leading signals.

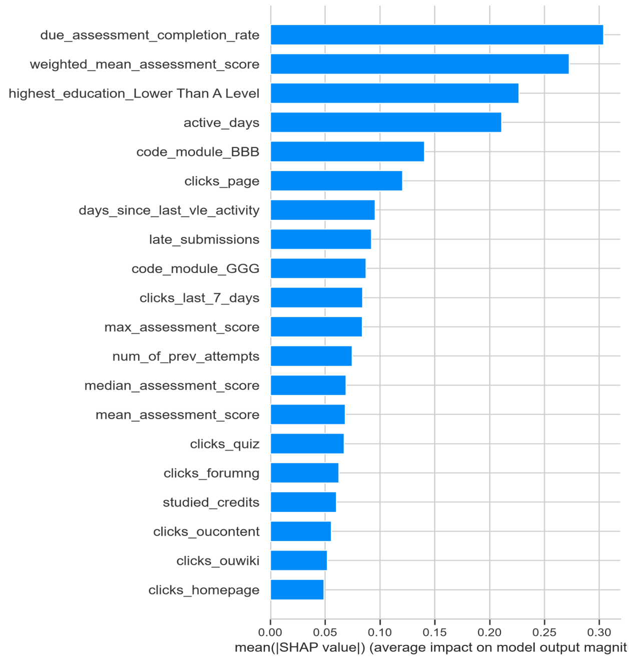

Figure 6. Global SHAP importance for the day-28 Gradient Boosting model

Figure 7 shows that lower completion and scores, fewer active days, longer inactivity, late submissions, and previous attempts generally shifted predictions toward risk. Because correlated engagement and score summaries can share attribution, these patterns explain the fitted model but should not be interpreted as causal or institution-invariant effects.

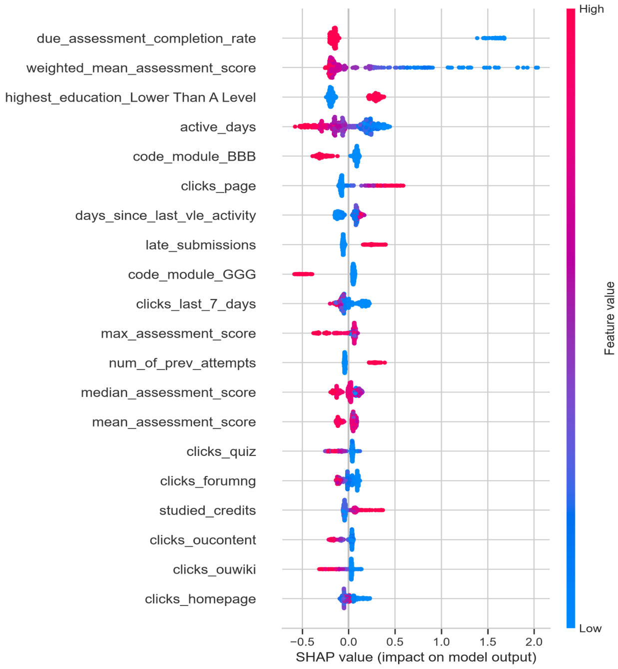

Figure 7. Direction and distribution of SHAP effects for leading predictors

4.7 Sensitivity and Subgroup Diagnostics

Table 7 summarizes the sensitivity specifications. Excluding demographic predictors reduced ROC-AUC from 0.790 to 0.778 and F1 from 0.656 to 0.642. Excluding assessment-score predictors reduced ROC-AUC to 0.762 and F1 to 0.629, showing that a more conservative score-free model retained useful but weaker discrimination. The three-class model achieved accuracy=0.636, balanced accuracy=0.490, and macro-F1=0.496. Recall was 0.903 for Success, 0.368 for Fail, and 0.199 for Withdrawn, confirming that the adverse outcomes are difficult to distinguish from each other using the available day-28 predictors. Figure 8 presents the three-class confusion matrix, showing that most errors involved Fail and Withdrawn cases being assigned to Success or to the other adverse-outcome category.

Table 7. Sensitivity analyses for alternative predictor and outcome specifications.

Specification	ROC-AUC	PR-AUC	F1
Full binary model	0.790	0.770	0.656
Demographics excluded	0.778	0.760	0.642
Assessment scores excluded	0.762	0.743	0.629
Three-class model	0.762*	—	0.496†

Weighted one-vs-rest ROC-AUC. †Macro-F1.

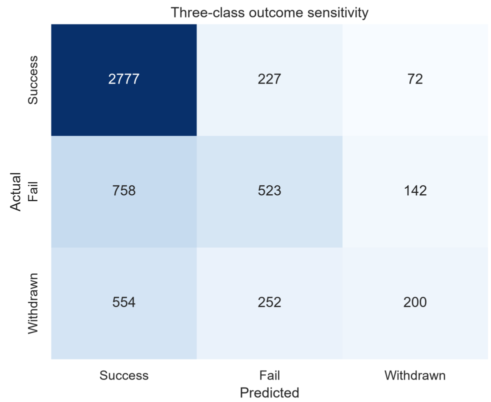

Figure 8. Three-class sensitivity analysis: Success, Fail, and Withdrawn.

Subgroup ROC-AUC was relatively stable for gender and disability groups, but threshold-dependent recall and false-positive rates differed. For example, recall was 0.549 for women and 0.636 for men; it was 0.534 for learners with A-level-equivalent prior education and 0.676 for those below A level. These differences partly reflect different risk prevalences and do not establish discrimination, but they preclude a claim that the model is demonstrably fair. Small subgroups, especially learners aged 55+ and some education categories, produced imprecise estimates.

5. Discussion

5.1 Principal Findings

The revised analysis supports a bounded claim: by day 28, OULAD records contain a meaningful but imperfect signal of later failure or withdrawal. Gradient Boosting provided the highest validation discrimination, yet its test advantage over Logistic Regression and Random Forest was only about 0.006 ROC-AUC. The key result is therefore not that a complex ensemble overwhelmingly outperformed all alternatives, but that several methods produced similar moderate discrimination when evaluated under the same leakage-controlled, student-grouped design.

The contrast with the original whole-dataset result is substantive. Once post-cutoff and unregistration-derived information was removed, accuracy declined from the previously reported 94% to 72%, and ROC-AUC from 0.985 to 0.790. This reduction is expected and strengthens credibility: performance now corresponds to information actually available at the declared prediction point.

5.2 Relationship to Prior OULAD Research

The findings align with prior evidence that prediction improves as the course progresses (Hlosta et al., 2017; Adnan et al., 2021; Zhang & Ahn, 2023). The day-28 ROC-AUC of 0.790 is lower than studies reporting 0.90 or above from later, whole-course, class-specific, or differently constructed tasks (Jha et al., 2019; Al-Zawqari et al., 2022; Ujkani et al., 2024). These values are not contradictions: they correspond to different targets, time points, feature sets, and validation schemes. The present design prioritizes a reproducible operational cutoff and student-disjoint evaluation over maximizing headline accuracy.

The SHAP results are also consistent with recent explainable OULAD studies in showing that VLE engagement, assessment behavior, prior attainment, and registration context carry substantial information (Zhang & Ahn, 2023; Ujkani et al., 2024). The revised model does not reproduce the earlier claim that a single last-activity variable explains approximately 85% of performance; instead, importance is distributed across assessment completion, scores, prior education, activity regularity, recency, and module context.

5.3 Interpretation of Failure and Withdrawal

Combining Fail and Withdrawn remains defensible for first-stage supportive triage because both outcomes may justify contact, and the three-class model had low recall for both adverse classes—especially withdrawal. Nevertheless, they should not be treated as pedagogically identical. Failure may reflect difficulty mastering assessed content, whereas withdrawal can reflect financial, health, employment, caring, motivational, or administrative factors not recorded in OULAD. Institutions should therefore follow a binary alert with human-led diagnosis rather than assigning a standardized intervention based solely on the model.

5.4 Practical and Managerial Implications

A day-28 model may help allocate limited advising capacity by ranking still-enrolled students for supportive outreach. The calibrated probabilities are more appropriate for prioritization than a rigid label. At the default threshold, recall was only 0.597, so institutions seeking to miss fewer at-risk learners would need a lower threshold and should anticipate more false positives. Thresholds should be chosen according to intervention capacity, the cost of unnecessary contact, and the harm of missed support.

The SHAP findings suggest actionable monitoring domains rather than deterministic rules: completion of assessments already due, assessment performance when available, recent and sustained VLE engagement, late submission, and prior attempts. An alert should prompt a conversation and offer of assistance. It should not restrict access, reduce opportunities, or be interpreted as evidence of low ability.

5.5 Ethical and Equity Considerations

The analysis does not establish algorithmic fairness. Prior education was among the leading predictors, and subgroup sensitivity and false-positive rates differed at the 0.50 threshold. Removing demographic predictors produced only a modest performance decline, but omitting protected attributes does not remove structural inequality from behavioral traces. Deployment should include subgroup monitoring, transparent communication, appeal and correction mechanisms, data minimization, human review, and periodic recalibration. Outreach should be supportive and voluntary wherever possible.

5.6 Limitations and Future Research

First, OULAD represents one UK distance-learning institution in 2013-2014; the anonymized modules do not reveal programme or academic level. The dataset remains valuable as a stable, publicly documented benchmark for temporal reconstruction and direct methodological comparison, but it cannot represent post-pandemic platforms, current instructional designs, or contemporary learner behaviour. External validity to other institutions and current learner populations is therefore unknown, and the fitted model should not be deployed without prospective validation and recalibration on recent local data. Second, the operational cohort changes across prediction points because students who have already withdrawn are no longer eligible; the earliness curve should not be interpreted as a pure within-cohort experiment. Third, OULAD provides submission dates but not score-release timestamps, making the score-free sensitivity analysis important. Fourth, module and presentation indicators were influential, and course-presentation holdout performance was lower than ordinary student-grouped validation. Fifth, SHAP explanations are associational and can distribute importance across correlated variables. Sixth, subgroup estimates are descriptive and underpowered for small categories. Finally, the study evaluates prediction, not whether an intervention improves outcomes.

Future work should prospectively validate the protocol using recent multi-institutional learning-management-system data, compare fixed cohorts across prediction times, record when grades become operationally available, tune decision thresholds to intervention capacity, distinguish pedagogical from non-academic withdrawal mechanisms, and test whether human-led interventions produce benefits without disproportionate burden or stigma.

6. Conclusion

Using leakage-controlled information available by day 28, student-grouped validation, seven tuned classifiers, calibration, uncertainty estimates, robustness checks, and SHAP explanations, this study found moderate and operationally relevant prediction of later failure or withdrawal. Gradient Boosting achieved ROC-AUC=0.790 and PR-AUC=0.770 on a student-disjoint test set but recall at the default threshold was 0.597 and performance was close to several simpler alternatives. Assessment completion and performance, prior education, active days, and VLE recency were influential, while a score-free model retained lower but meaningful discrimination. The model should therefore be understood as a prioritization aid for supportive human outreach, not an autonomous decision system. The credible contribution is a transparent account of what can be predicted at a specified early point and of the uncertainty, heterogeneity, and governance constraints that accompany that prediction.

Declarations

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Ethical Approval: Ethics committee approval was not required for this study because the analyses were conducted exclusively using the publicly available, secondary, and anonymized Open University Learning Analytics Dataset (OULAD). The authors declare that all scientific and ethical principles were observed throughout the conduct and preparation of the study.

Author Contributions: Conceptualization, V.G.O. (50%) and Y.S.B. (50%); Data Curation, V.G.O. (50%) and Y.S.B. (50%); Investigation, Analysis and Validation, V.G.O. (50%) and Y.S.B. (50%); Writing – Original Draft Preparation, V.G.O. (50%) and Y.S.B. (50%); Writing – Review and Editing, V.G.O. (50%) and Y.S.B. (50%). All authors have read and approved the final version of the manuscript.

Data Availability: The data used in this study are publicly available through the Open University Learning Analytics Dataset (OULAD). No new primary data were collected for this study.

Informed Consent: Not applicable. The study was based exclusively on publicly available, secondary, and anonymized data and did not involve direct recruitment of or interaction with human participants.

AI Disclosure: The scholarly content of this study, including the research questions, theoretical framework, methodology, analyses, interpretation, and conclusions, was produced by the authors. During the preparation of the manuscript, the AI-based tool Deeply was used to a limited extent solely for language editing, improving the fluency of expression, and formal checks. All AI-assisted outputs were carefully reviewed and verified by the authors. AI-based tools were not used for data collection, data generation, statistical analysis, or citation/reference generation. The authors take full responsibility for the final content of the manuscript.

References

Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., Bashir, M., & Khan, S. U. (2021). Predicting at-risk students at different percentages of course length for early intervention using machine learning models. IEEE Access, 9, 7519–7539. https://doi.org/10.1109/ACCESS.2021.3049446

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. Telematics and Informatics, 37, 13–49. https://doi.org/10.1016/j.tele.2019.01.007

Al-Zawqari, A., Peumans, D., & Vandersteen, G. (2022). A flexible feature selection approach for predicting students' academic performance in online courses. Computers and Education: Artificial Intelligence, 3, 100103. https://doi.org/10.1016/j.caeai.2022.100103

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 267–270). ACM. 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

Basilaia, G., & Kvavadze, D. (2020). Transition to online education in schools during a SARS-CoV-2 coronavirus (COVID-19) pandemic in Georgia. Pedagogical Research, 5(4), Article em0060. https://doi.org/10.29333/pr/7937

Bravo, L.E.C., Lopez, H. J. F., & Pimiento, N. N., Educational data integration and machine learning for academic performance prediction. Ingeniería Solidaria, 21 (2). https://doi.org/10.16925/2357-6014.2025.02.07

Chen, F., & Cui, Y. (2023). Utilizing student time series behavior in learning management systems for early prediction of course performance. Journal of Learning Analytics, 10(1), 1–18. https://doi.org/10.18608/jla.2023.7397

Dabbagh, N., & Fake, H. (2017). College students' perceptions of personal learning environments through the lens of digital tools, processes and spaces. Journal of New Approaches in Educational Research, 6(1), 28–36. https://doi.org/10.7821/naer.2017.1.215

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue. In Proceedings of the 6th International Conference on Learning Analytics and Knowledge (pp. 89–98). ACM. https://doi.org/10.1145/2883851.2883893

Essa, A., & Ayad, H. (2012). Student success system: Risk analytics and data visualization using ensembles of predictive models. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 158–161). ACM. https://doi.org/10.1145/2330601.2330641

European Commission. (2020). European skills agenda for sustainable competitiveness, social fairness and resilience. Publications Office of the European Union.

Ferguson, R. (2013). 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

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

Gasevic, D., Kovanovic, V., Joksimovic, S., & Siemens, G. (2014). Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. International Review of Research in Open and Distributed Learning, 15(5), 134–176. https://doi.org/10.19173/irrodl.v15i5.1954

Herodotou, C., Maguire, C., Hlostra, M., & Mulholland, P., (2023). Predictive Learning Analytics and University Teachers: Usage and perceptions three years post implementation. Conference: LAK 2023: 13th International Learning Analytics and Knowledge Conference 68-78. 10.1145/3576050.3576061

Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. In Proceedings of the 7th International Learning Analytics and Knowledge Conference (pp. 6–15). ACM. https://doi.org/10.1145/3027385.3027449

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. EDUCAUSE Review. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide agenda. International Journal of Artificial Intelligence in Education, 32(2), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Huptych, M., Hlosta M., Zdrahal Z., Kocvara J., (2018) Investigating Influence of Demographic Factors on Study Recommenders Artificial Intelligence in Education https://doi.org/10.1007/978-3-319-93846-2_27

Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. Computational Intelligence and Neuroscience, 2018, Article 6347186. https://doi.org/10.1155/2018/6347186

Jha, N. I., Ghergulescu, I., & Moldovan, A.-N. (2019). OULAD MOOC dropout and result prediction using ensemble, deep learning and regression techniques. In Proceedings of the 11th International Conference on Computer Supported Education (pp. 154–164). https://doi.org/10.5220/0007767901540164

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. International Review of Research in Open and Distributed Learning, 15(1), 133–160. https://doi.org/10.19173/irrodl.v15i1.1651

Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades. Artificial Intelligence Review, 37(4), 331–344. https://doi.org/10.1007/s10462-011-9234-x

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics Dataset. Scientific Data, 4, Article 170171.

https://doi.org/10.1038/sdata.2017.171

Law, N., & Liang, L. (2020). A Multilevel Framework and Method for Learning Analytics Integrated Learning Design. Journal of Learning Analytics, 7(3), 98-117. https://doi.org/10.18608/jla.2020.73.8

Lim, L. A., Dawson, S., Gasevic, D., Joksimovic, S., Fudge, A., Pardo, A., & Gentili, S. (2022). Students' sense-making of personalised feedback based on learning analytics. Australasian Journal of Educational Technology, 38(4), 15–34. https://doi.org/10.14742/ajet.6370

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774. 10.48550/arXiv.1705.07874

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

Mduma, N., Kalegele, K., & Machuve, D. (2019). A survey of machine learning approaches and techniques for student dropout prediction. Data Science Journal, 18(1), 14. https://doi.org/10.5334/dsj-2019-014

Merriam, S. B., & Bierema, L. L. (2014). Adult learning: Linking theory and practice. Jossey-Bass.

Onah, D. F. O., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: Behavioural patterns. In Proceedings of the 6th International Conference on Education and New Learning Technologies (EDULEARN14) (pp. 5825–5834). IATED. 10.13140/RG.2.1.2402.0009

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Educational Technology & Society, 17(4), 49–64.

Prenkaj, B., Velardi, P., Stilo, G., Distante, D., & Faralli, S. (2020). A survey of machine learning approaches for student dropout prediction in online courses. ACM Computing Surveys, 53(3), Article 62. https://doi.org/10.1145/3388792

Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: The obligation to act. In Proceedings of the 7th International Learning Analytics and Knowledge Conference (pp. 46–55). ACM. https://doi.org/10.1145/3027385.3027406

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. Computers in Human Behavior, 60, 333–341. https://doi.org/10.1016/j.chb.2016.02.074

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

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. Computer Applications in Engineering Education, 21(1), 135–146. https://doi.org/10.1002/cae.20456

Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education: A review of UK and international practice. JISC.https://www.academia.edu/41021774/Learning_Analytics_in_Higher_Education_A_review_of_UK_and_international_practice_Full_report

Selwyn, N. (2019). What is digital sociology? Polity Press.

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

Tsai, Y. S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., Drachsler, H., Delgado Kloos, C., & Gasevic, D. (2020). Learning analytics in European higher education — Trends and barriers. Computers & Education, 155, Article 103933. https://doi.org/10.1016/j.compedu.2020.103933

Ujkani, B., Minkovska, D., & Hinov, N. (2024). Course success prediction and early identification of at-risk students using explainable artificial intelligence. Electronics, 13(21), 4157. https://doi.org/10.3390/electronics13214157

UNESCO. (2016). Education 2030: Incheon Declaration and framework for action. UNESCO.

Viberg, O., Khalil, M., & Baars, M. (2020). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. In Proceedings of the 14th International Conference on Learning Analytics and Knowledge (pp. 524–533). ACM. https://doi.org/10.1145/3375462.3375483

Wei, X., Saab, N., Admiraal, W., (2023). Do learners share the same perceived learning outcomes in MOOCs? Identifying the role of motivation, perceived learning support, learning engagement, and self-regulated learning strategies. The Internet and Higher Education, 53. https://doi.org/10.1016/j.iheduc.2022.100880

Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in higher education: Critical issues and perspectives. Teaching in Higher Education, 25(4), 351–365. https://doi.org/10.1080/13562517.2020.1748811

Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: Digital technologies and distance education during the coronavirus emergency. Learning, Media and Technology, 45(2), 107–114. https://doi.org/10.1080/17439884.2020.1761641

Xu, J., Moon, K. H., & van der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. IEEE Journal of Selected Topics in Signal Processing, 11(5), 742–753. https://doi.org/10.1109/JSTSP.2017.2692560

Zawacki-Richter, O., & Naidu, S. (2016). Mapping research trends from 35 years of publications in Distance Education. Distance Education, 37(3), 245–269. https://doi.org/10.1080/01587919.2016.1185079

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, 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), Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, C., & Ahn, H. (2023). E-learning at-risk group prediction considering the semester and realistic factors. Education Sciences, 13(11), 1130. https://doi.org/10.3390/educsci13111130

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary Educational Psychology, 25(1), 82–91. https://doi.org/10.1006/ceps.1999.1016

Appendix A. Representative Analytical Records

Table A1 presents ten genuine records selected reproducibly from the day-28 analytical dataset (five at-risk and five not-at-risk enrolments). The original anonymized student identifier was removed and replaced by a sequential record code. Combined cells report related measures compactly; a dash indicates that no applicable assessment score was observable by day 28

Table A1. De-identified representative records from the day-28 analytical dataset

Record	Module / term	Sex / age	Prior education	Clicks / days	VLE recency	Completion / score	Outcome / risk
R001	BBB / 2014J	F / 35-55	A-level equivalent	312 / 11	3	1.00 / 1.0	Pass / Not at risk
R002	CCC / 2014J	M / 0-35	A-level equivalent	650 / 17	0	1.00 / 100.0	Distinction / Not at risk
R003	FFF / 2014B	M / 35-55	Below A-level	453 / 20	1	1.00 / 78.0	Pass / Not at risk
R004	CCC / 2014B	M / 35-55	A-level equivalent	708 / 24	1	1.00 / 70.0	Distinction / Not at risk
R005	FFF / 2014J	M / 35-55	Below A-level	1,374 / 30	0	1.00 / 72.0	Pass / Not at risk
R006	BBB / 2014J	M / 35-55	Below A-level	277 / 9	2	1.00 / 1.0	Withdrawn / At risk
R007	DDD / 2013B	M / 0-35	Below A-level	354 / 22	0	1.00 / 76.6	Withdrawn / At risk
R008	EEE / 2014B	M / 0-35	A-level equivalent	171 / 8	5	- / -	Fail / At risk
R009	DDD / 2013J	F / 0-35	A-level equivalent	105 / 12	4	1.00 / 50.0	Fail / At risk
R010	FFF / 2014J	F / 0-35	HE credential	253 / 23	2	1.00 / 60.0	Withdrawn / At risk