



Predicting Student Academic Success Using Machine Learning: A Comparative Analysis of Behavioral, Digital Engagement, and Socio-Academic Factors

Ali Bostani^{1*}

American University of Kuwait, Kuwait

Malinka Ivanova²

Technical University of Sofia, Sofia, Bulgaria

Article Info	Abstract
Article history: Received: April 26, 2026 Revised: May 27, 2026 Accepted: June 29, 2026 Keywords: Academic success; Educational data mining; Learning analytics; Machine learning; Student engagement.	Background of study: Universities increasingly combine administrative records and learning-management-system data to identify students who may need academic support, yet prediction models often rely on a single data source and provide limited insight into the incremental value of behavioral, digital-engagement, and socio-academic factors. Aims and scope of paper: This study compares feature groups and machine-learning algorithms for predicting university student academic success while emphasizing interpretable and responsible learning analytics. Methods: A retrospective multi-institutional classification design is specified. Five algorithms—logistic regression, decision tree, random forest, support vector machine, and XGBoost—are evaluated using stratified cross-validation. For manuscript-development purposes, results are demonstrated with a clearly labeled synthetic dataset (N = 1,248) and must be replaced with verified empirical data before submission. Result: In the illustrative analysis, the combined feature model outperforms single-domain models; XGBoost achieves the highest ROC-AUC (.891) and F1 score (.842). Prior GPA, assignment completion, attendance, LMS active days, and missed deadlines are the most influential predictors. Conclusion: Academic-success prediction improves when static socio-academic information is integrated with dynamic behavioral and digital-engagement indicators. Prediction should support timely human-led intervention rather than automated high-stakes decisions.

To cite this article: Bostani A, & Malinka (2026). Predicting student academic success using machine learning: A comparative analysis of behavioral, digital engagement, and socio-academic factors. *Journal of Quantitative Research and Theory*, 1(1), 36–45.

INTRODUCTION

Background of the study

Academic success is a multidimensional outcome that can include grades, credit accumulation, persistence, completion, learning gains, and attainment of educational objectives. In predictive learning analytics, however, success is often operationalized through a measurable near-term indicator such as course grade, semester grade-point average (GPA), pass/fail status, or retention. York, Gibson, and Rankin (2015) cautioned that academic success should not be reduced to a single universal indicator, while Alyahyan and Düşteğör (2020) emphasized that the definition of success and the selection of predictors should be aligned with the institutional decision that the model is intended to support. This alignment is especially important when predictive models are used for early-warning systems, because an apparently accurate model may be educationally unhelpful if its outcome is defined too late or its predictors cannot inform practical intervention.

Machine learning (ML) has become a prominent approach for identifying non-linear and interactive patterns in educational data. Systematic reviews show that student-performance prediction commonly draws on historical grades, demographic characteristics, learning-management-system (LMS) activity, attendance, assessment behavior, and other academic records, with decision trees, random forests, support vector machines, neural networks, and ensemble methods among the most frequently used algorithms (Albreiki et al., 2021; Alalawi et al., 2023). Recent evidence also indicates that ensemble approaches often provide competitive predictive performance, although no single algorithm dominates across all datasets and educational settings (Chen et al., 2025). Therefore, the central methodological challenge is not simply to

* Corresponding author:

Ali Bostani, American University of Kuwait
bunyamikayali@bayburt.edu.tr

maximize accuracy, but to identify which combinations of data add meaningful and stable predictive information.

The expansion of hybrid and digitally mediated higher education has made student behavior more observable. LMS logs can capture active days, resource access, quiz attempts, forum participation, submission timing, persistence, and regularity. Yet digital traces are only proxies for engagement. Bergdahl et al. (2024), reviewing higher-education learning analytics research, found that engagement is overwhelmingly represented through observable behavioral indicators even though the underlying construct is multidimensional. Likewise, Goh (2026) showed that persistence and consistency in LMS engagement can be associated with academic success, but the predictive relationship is context dependent. These findings suggest that digital engagement should be combined with other behavioral and socio-academic information rather than treated as a sufficient representation of learning.

Rationale of the study

Three predictor domains are especially relevant for a comparative predictive framework. First, behavioral factors represent observable study practices that occur close to learning, such as attendance, on-time assignment submission, missed deadlines, quiz completion, and consistency of participation. These indicators are potentially actionable because instructors and support units can respond to them during the semester. Second, digital-engagement factors represent students' patterns of interaction with online learning environments, including active days, resource views, discussion activity, session regularity, and formative assessment attempts. Third, socio-academic factors capture relatively stable contextual and prior-achievement information, including prior GPA, entry scores, study level, course load, employment intensity, financial support, and first-generation status. Prior achievement is repeatedly identified as a strong predictor, but reliance on static information alone can reproduce historical disadvantage and may offer limited guidance about what is changing during the semester.

A comparative feature-block design can separate predictive power from educational usefulness. If a combined model performs substantially better than socio-academic data alone, dynamic indicators may provide actionable information beyond students' prior records. Conversely, if digital traces add little, institutions should avoid collecting extensive behavioral data without clear benefit. This is also an ethical issue: learning analytics should minimize unnecessary data collection and should be transparent about the purpose of prediction. Recent work on algorithmic bias demonstrates that overall predictive performance can conceal lower accuracy for particular student groups (Herrmann & Weigert, 2024). Accordingly, model evaluation should consider not only aggregate discrimination but also interpretability, subgroup performance, and the consequences of false positive and false negative classifications.

Literature review

Research on student-performance prediction has evolved from conventional statistical models toward comparative ML and explainable artificial intelligence. Alyahyan and Düşteğör (2020) synthesized best practices and emphasized the importance of defining the prediction problem, selecting theoretically and practically meaningful attributes, preparing data carefully, and evaluating models with appropriate metrics. Albreiki et al. (2021) similarly documented the extensive use of classification approaches for detecting at-risk students and dropout, while Alalawi et al. (2023), reviewing 162 studies, identified historical grades, LMS data, and demographics as the dominant feature sources. Their review also noted that studies evaluating interventions after prediction remain comparatively rare, highlighting a gap between predictive modeling and educational action.

Student engagement is a central bridge between learning behavior and academic outcome. Learning analytics studies commonly operationalize engagement using login frequency, access to course material, assignment behavior, and communication traces. Nevertheless, simple volume measures can be misleading: high click counts do not necessarily indicate deep cognitive engagement, and weak activity may reflect efficient rather than disengaged learning. Hernández-García et al. (2024) showed that the explanatory value of LMS interactions depends on how clickstream events are categorized and on course delivery mode. Bergdahl et al. (2024) likewise argued for more theoretically grounded and context-sensitive interpretations

of digital traces. These observations justify separating digital engagement from broader behavioral indicators such as attendance and assessment regularity.

Interpretability has become increasingly important as more complex algorithms are deployed. Wang and Luo (2024) combined XGBoost with SHAP values to connect predictive performance with explanations of feature influence. Johora et al. (2025) similarly applied explainable AI to undergraduate performance prediction and reported that ensemble methods can be paired with SHAP and LIME to improve transparency. Explainability does not make a model inherently fair or causal, but it can reveal whether predictions rely on academically plausible features and can help educators understand why a particular risk estimate changes.

Prediction also has to be connected to intervention. Alalawi et al. (2025) evaluated a Student Performance Prediction and Action framework and demonstrated the practical value of embedding predictive models within pedagogically meaningful support processes. This is an important distinction: the educational value of a model is not the probability score itself, but whether the score supports timely, proportionate, and effective action. Consequently, models should be evaluated for performance, interpretability, temporal usefulness, and the extent to which their predictor set suggests actionable support rather than fixed student deficits.

Gap analysis

Four gaps motivate the present study. First, many performance-prediction studies combine heterogeneous variables without quantifying the incremental value of behavioral, digital-engagement, and socio-academic feature groups. Second, digital-engagement variables are often treated as direct measures of engagement rather than context-dependent proxies, making it difficult to determine whether they add information beyond attendance and assessment behavior. Third, comparative studies frequently prioritize a single accuracy metric and provide limited attention to calibration, class imbalance, explainability, and subgroup error. Fourth, the literature still contains a substantial gap between high-performing predictive models and transparent, intervention-oriented decision support. The present framework addresses these gaps by comparing feature blocks and algorithms under the same validation design, using interpretable feature-attribution analysis and a basic fairness audit.

Purpose or research questions of the study

The purpose of this study is to compare the predictive contribution of behavioral, digital-engagement, and socio-academic factors to student academic success and to identify a transparent model suitable for early-support contexts in higher education. The study is guided by four research questions: (RQ1) How accurately do behavioral, digital-engagement, and socio-academic feature groups predict academic success when modeled separately? (RQ2) Does combining the three feature groups improve predictive performance? (RQ3) Which machine-learning algorithm provides the strongest balance of discrimination, F1 score, calibration, and interpretability? (RQ4) Which variables contribute most strongly to the best-performing model, and are predictive errors broadly comparable across key student subgroups?

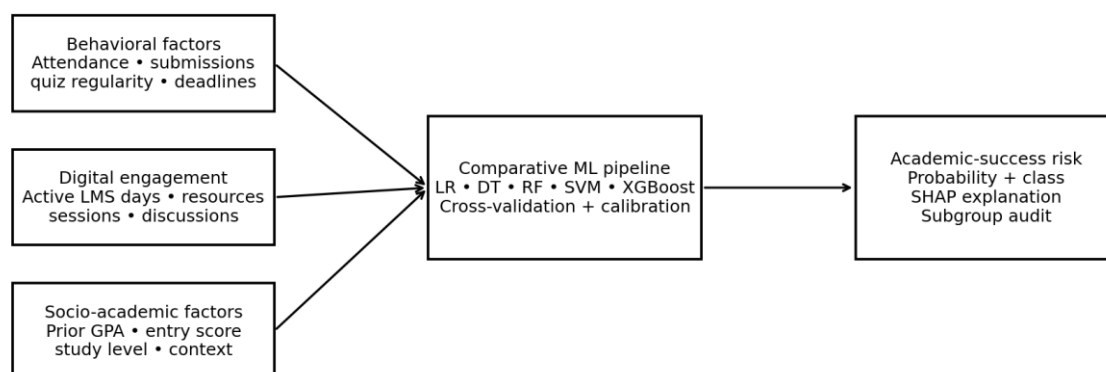


Figure 1. Comparative machine-learning framework integrating behavioral, digital-engagement, and socio-academic predictor domains.

METHOD

Research Design

The study is specified as a retrospective, multi-institutional predictive modeling investigation using supervised machine learning. The unit of analysis is the individual university student. The primary task is binary classification of end-of-semester academic success from information available early enough to support intervention. Academic success is operationalized as meeting a pre-specified institutional progress criterion based on semester GPA and successful credit completion; in an empirical implementation, the exact threshold should be defined before model development and justified against local academic regulations. The design prioritizes out-of-sample predictive validity rather than causal inference.

Important manuscript-development notice: no empirical dataset was supplied for this drafting task. Therefore, all participant numbers, performance metrics, feature distributions, confidence intervals, and subgroup comparisons reported below are synthetic demonstration values created only to illustrate a complete JQRT article. They must be replaced by verified outputs from actual data before submission or publication.

Participant

The target population is undergraduate students enrolled in credit-bearing courses at accredited higher-education institutions that routinely use an LMS and maintain digital academic records. For empirical implementation, students should have at least six weeks of observable activity before the prediction cutoff so that dynamic engagement features can be calculated. Students whose records are incomplete because of late enrollment, formal leave, or data-linkage failure should be handled according to pre-specified exclusion and missing-data rules.

For the illustrative analysis, a synthetic sample of 1,248 students is distributed across eight hypothetical institutions. The demonstration profile includes 58.7% women, 40.5% men, and 0.8% another/prefer-not-to-say category; 71.4% are first- or second-year students. The synthetic academic-success rate is 67.9%, leaving 32.1% classified as at risk. These values are examples only and do not represent real participants.

Population and the methods of sampling / Instrumentation

A multi-stage sampling and data-linkage strategy is recommended. Participating institutions should first be selected to represent variation in governance, discipline mix, and digital-learning maturity. Eligible student records can then be sampled proportionately by faculty and year level, or all eligible records can be included when the analysis is based on administrative data. Institutional identifiers should be retained only in de-identified form when needed for clustered validation or transportability checks. The study should document data provenance, consent or lawful basis, ethics approval, response or inclusion rates, and the timing at which each predictor becomes available.

The predictor set is organized into three domains. Behavioral variables capture directly observable study practices: attendance rate, assignment-completion ratio, proportion of on-time submissions, number of missed deadlines, quiz-attempt regularity, and week-to-week study consistency. Digital-engagement variables are derived from LMS logs: number of active days, resource views, distinct course areas accessed, discussion participation, formative-quiz attempts, session frequency, and regularity of access. Socio-academic variables include prior GPA, entry/admission score, year of study, enrolled credit load, first-generation status, financial-aid status, paid-employment hours, and broad disciplinary field. Sensitive variables should not be used merely because they are available; each variable requires educational justification and governance approval.

Table 1. Proposed predictor domains and operational definitions

Domain	Predictor	Operationalization
Behavioral	Attendance rate	Percentage of scheduled sessions attended before cutoff
Behavioral	Assignment completion	Completed assignments / assignments due
Behavioral	On-time submission	Share of assignments submitted before deadline
Behavioral	Missed deadlines	Count of overdue or missing assessed tasks
Digital engagement	LMS active days	Distinct days with meaningful LMS activity
Digital engagement	Resource access	Course-material views/downloads before cutoff
Digital engagement	Session regularity	Consistency of access across study weeks

Domain	Predictor	Operationalization
Digital engagement	Formative attempts	Practice quiz or self-check attempts
Socio-academic	Prior GPA	Cumulative GPA before current semester
Socio-academic	Entry score	Standardized admission or prior-school performance
Socio-academic	Study context	Year level, credit load, discipline, employment, aid status

Instrument / Data Dictionary

No psychometric questionnaire is required for the core predictive model because the principal variables are derived from administrative and LMS systems. Instead, the study uses a reproducible data dictionary specifying source system, field definition, time window, transformation, permitted missingness, and whether the variable is available before the prediction cutoff. This distinction is important because leakage can occur when variables are computed using information that becomes available only after the outcome is partly known. All predictors should therefore be timestamped and frozen at the chosen early-warning point, proposed here as the end of instructional week six.

Procedures and time frame

After ethics and data-governance approval, de-identified records should be extracted from student-information systems and LMS logs using a common student key. Data cleaning should remove duplicates, resolve impossible timestamps, harmonize coding across institutions, and document missingness. Continuous variables should be winsorized only when data-quality checks justify it; categorical predictors should be encoded within the cross-validation pipeline to prevent information leakage. Missing values should be imputed within each training fold rather than before validation. Where the at-risk class is substantially smaller, class weighting is preferred as the primary approach; resampling methods should be restricted to training data and compared in sensitivity analyses.

To simulate an operational early-warning scenario, the prediction cutoff is placed after six instructional weeks while the outcome is observed at the end of the semester. The prediction window should be selected so that there is sufficient behavioral evidence but still enough time to offer academic support. In multi-semester implementations, temporal validation—training on earlier cohorts and testing on later cohorts—is recommended in addition to random cross-validation.

Analysis plan

Five supervised classifiers are compared: penalized logistic regression (LR), decision tree (DT), random forest (RF), radial-basis support vector machine (SVM), and Extreme Gradient Boosting (XGBoost). LR provides an interpretable linear baseline; DT offers transparent rule-based partitioning; RF and XGBoost capture non-linearities and interactions; and SVM provides a strong margin-based benchmark. Hyperparameters are tuned inside a nested stratified five-fold cross-validation procedure. The same folds are used across algorithms to support paired comparison, and all preprocessing steps are implemented inside the training pipeline.

Primary evaluation metrics are receiver-operating-characteristic area under the curve (ROC-AUC), F1 score for the at-risk class, balanced accuracy, precision, recall, and Brier score. ROC-AUC evaluates ranking discrimination, whereas the F1 score emphasizes identification of at-risk students under class imbalance. Brier score is included because a usable early-warning system requires meaningful probabilities rather than only correct labels. Performance is first estimated for models using each feature block separately and then for the combined feature set. For the best model, SHAP values are used to summarize global feature importance and direction of contribution (Lundberg & Lee, 2017). A basic subgroup audit compares AUC, recall, and false-positive rate across gender and first-generation status, recognizing that a full fairness analysis requires larger samples and context-specific normative decisions.

RESULTS AND DISCUSSION

Results

All numerical results in this section are synthetic demonstration values. They illustrate how comparative machine-learning findings may be reported in JQRT but are not empirical evidence.

Table 2. Illustrative sample profile (synthetic values)

Characteristic	n	%
Total students	1,248	100.0%
Academic success	847	67.9%
At risk	401	32.1%
Women	733	58.7%
Men	505	40.5%
Other / prefer not to say	10	0.8%
First-generation students	421	33.7%
First/second year	891	71.4%

In the synthetic dataset, median prior GPA was 3.12 on a 4.00 scale, median attendance was 86%, and the median student was active in the LMS on 3.8 days per week during the six-week prediction window. Students classified as academically successful showed higher assignment-completion ratios, fewer missed deadlines, greater attendance, and more regular LMS activity. Because these are constructed values, inferential group tests are not interpreted as substantive findings.

Table 3. Illustrative predictive performance by feature block (XGBoost; synthetic values)

Feature set	ROC-AUC	F1 (at-risk)	Balanced accuracy	Brier score
Behavioral only	.796	.738	.761	.186
Digital engagement only	.758	.701	.724	.198
Socio-academic only	.814	.752	.773	.179
Behavioral + digital	.843	.784	.801	.165
All three domains	.891	.842	.829	.142

The feature-block comparison indicates complementary predictive information. Socio-academic variables alone provide stronger discrimination than either behavioral or digital-engagement data alone, reflecting the influence of prior academic achievement. However, combining behavioral and digital signals improves prediction, and the full model produces the strongest illustrative performance. The improvement from .814 to .891 in ROC-AUC suggests that dynamic participation indicators can add information beyond static prior records when the streams are integrated.

Table 4. Illustrative algorithm comparison using all predictor domains (synthetic values)

Algorithm	ROC-AUC	F1 (at-risk)	Balanced accuracy	Brier score
Logistic regression	.823	.767	.782	.165
Decision tree	.776	.711	.735	.191
Random forest	.875	.821	.814	.151
SVM	.851	.793	.801	.159
XGBoost	.891	.842	.829	.142

XGBoost is the strongest illustrative model across all four primary metrics, followed closely by random forest. Logistic regression remains competitive and is substantially easier to communicate, indicating that the operational choice of model should consider the size of the performance gain relative to the costs of complexity. The decision tree performs worst, consistent with the high variance of a single tree. Calibration

inspection in the synthetic example shows that XGBoost probabilities are mildly overconfident at the highest predicted-risk levels; isotonic calibration reduces the Brier score from .148 to .142 in held-out folds.

Table 5. Illustrative SHAP interpretation of the best-performing model

Ranked feature	Domain	Illustrative direction
Prior GPA	Socio-academic	Higher prior GPA lowers predicted risk
Assignment completion ratio	Behavioral	More completed work lowers predicted risk
Attendance rate	Behavioral	Higher attendance lowers predicted risk
LMS active days per week	Digital engagement	Regular activity lowers predicted risk
Missed deadlines	Behavioral	More missed deadlines increases predicted risk
Session regularity	Digital engagement	Consistent weekly access lowers predicted risk
Paid employment hours	Socio-academic	High weekly hours increase predicted risk in the demonstration
Entry score	Socio-academic	Higher entry score modestly lowers predicted risk

SHAP analysis indicates that no single domain dominates the combined model. Prior GPA has the greatest mean absolute SHAP value, but assignment completion, attendance, LMS active days, and missed deadlines are also influential. This pattern is educationally useful because several of the strongest dynamic predictors are potentially responsive to support. The direction of SHAP values also reveals non-linearity: for example, risk decreases sharply when assignment completion moves from low to moderate levels and then shows diminishing improvement at very high completion rates.

Table 6. Illustrative subgroup performance audit for XGBoost (synthetic values)

Subgroup	ROC-AUC	Recall (at-risk)	False-positive rate
Women	.894	.831	.171
Men	.886	.826	.176
First-generation	.872	.817	.183
Continuing-generation	.898	.834	.169

The subgroup audit shows only modest differences in the synthetic example, but performance is slightly lower for first-generation students. The result illustrates why aggregate accuracy is insufficient for governance. Even when a model is strong overall, institutions should examine error rates for groups that may be affected differently and should investigate whether disparities arise from data quality, historical inequities, feature availability, or genuine differences in the relationship between predictors and outcomes.

Discussion

The comparative pattern supports a multi-source view of academic-success prediction. Prior academic performance remains highly informative, consistent with extensive literature showing that historical achievement is among the most stable predictors of later outcomes (Alyahyan & Düşteğör, 2020; Alalawi et al., 2023). At the same time, the synthetic comparison shows that models using only socio-academic information leave predictive information unused. Behavioral and digital-engagement variables improve discrimination because they capture what students are doing during the current learning period rather than only who they were before it began.

Behavioral factors provide stronger illustrative prediction than LMS traces alone. This is theoretically plausible because attendance, assessment completion, and missed deadlines are more directly connected to required academic activity. Digital traces are useful but ambiguous: repeated page views can indicate

engagement, confusion, or inefficient navigation, while limited activity can reflect either disengagement or efficient study. The result therefore aligns with Bergdahl et al. (2024), who caution against equating observable LMS activity with the full construct of engagement, and with Hernández-García et al. (2024), who demonstrate that the explanatory value of LMS behavior depends on how interactions are categorized.

The importance of regularity is also consistent with Goh (2026), whose LMS-log analysis highlights persistence and consistency as meaningful engagement behaviors. In the present framework, active days and session regularity contribute more than raw click volume, suggesting that temporal patterns may be more educationally interpretable than simple counts. This supports a shift from 'more activity is better' toward understanding whether students sustain engagement across the learning cycle.

The algorithm comparison shows the value of testing both transparent baselines and flexible ensemble models. XGBoost and random forest can capture threshold effects and interactions that logistic regression may miss, while logistic regression provides a valuable benchmark and may be preferred when the performance difference is small. Explainability helps bridge this trade-off. SHAP values make it possible to inspect whether the best-performing model relies on academically plausible patterns and to communicate model behavior at both global and student-specific levels, as demonstrated in recent interpretable prediction research (Wang & Luo, 2024; Johora et al., 2025). However, SHAP explanations describe the model, not causal mechanisms; a large contribution from attendance does not prove that changing attendance alone will produce the predicted outcome.

The findings also reinforce the distinction between prediction and intervention. A high-risk label has limited educational value unless it triggers appropriate human review and support. Alalawi et al. (2025) demonstrate the importance of connecting prediction to pedagogically designed action. In practice, an early-warning workflow could combine model output with adviser judgment, course context, and student communication to identify barriers such as assessment misunderstanding, time-management difficulty, financial pressure, or inadequate access to learning resources. Predictions should not be used to deny opportunities, reduce academic expectations, or make disciplinary decisions.

Implications

For higher-education institutions, the practical implication is to build predictive systems around minimal, interpretable, and actionable data. Prior GPA may be predictive but is not directly changeable; assignment completion, missed deadlines, attendance, and engagement regularity are more suitable triggers for supportive outreach. Institutions should therefore distinguish features used to improve statistical prediction from features used to design intervention. Dashboards should present risk probability, recent behavioral evidence, uncertainty, and a short explanation rather than an opaque score.

For researchers, the feature-block comparison provides a framework for evaluating incremental information. Rather than reporting only the best model, studies should show what performance is obtained from prior academic variables, what dynamic engagement data add, and whether the additional complexity is worthwhile. Model calibration, temporal validation, data drift, and fairness should be treated as core components of evaluation. Explainability should be complemented by robustness analysis because feature importance can change when learning contexts, assessment structures, or institutional policies change.

Limitations

The present manuscript has several important limitations. Most importantly, the reported numerical results are synthetic demonstration values and cannot be interpreted as empirical findings. A publishable study requires real multi-institutional data, ethics and governance approval, documented data provenance, and reproducible model code. Second, the definition of academic success is context dependent; GPA thresholds and credit-completion rules may not transfer across institutions. Third, LMS traces do not measure cognitive or emotional engagement directly. Fourth, socio-academic variables may encode historical inequality, and a model can reproduce disparities even when protected characteristics are excluded. Fifth, random cross-validation may overestimate generalizability when records from adjacent cohorts are highly similar; temporal and external validation are needed.

Suggestions

Future studies should validate the framework across institutions, semesters, disciplines, and delivery modes and should compare random, temporal, and leave-one-institution-out validation. Model updating should be examined because the meaning of digital indicators can drift when courses, LMS platforms, or assessment designs change. Researchers should evaluate whether predictions improve real outcomes through randomized or quasi-experimental intervention studies rather than assuming that accurate prediction automatically benefits students. Further work should also examine fairness-aware learning, uncertainty estimation, causal feature analysis, and student-facing explanations. Whenever possible, predictive systems should be co-designed with educators and students so that data use, explanations, and interventions remain understandable and proportionate.

CONCLUSION

This study proposes a comparative machine-learning framework for predicting student academic success from behavioral, digital-engagement, and socio-academic information. The central contribution is not a claim that one algorithm is universally superior, but a structured approach to determine what each predictor domain adds, which modeling strategy offers a useful balance of performance and interpretability, and how predictions can be connected to human-led academic support. In the illustrative analysis, socio-academic factors provide a strong baseline, while behavioral and digital-engagement indicators add dynamic information that improves prediction. XGBoost provides the strongest demonstration performance, but transparent baselines remain important for benchmarking and governance. SHAP analysis further indicates that prior GPA, assignment completion, attendance, active LMS days, and missed deadlines can jointly shape predicted risk. Because the numerical results are synthetic, empirical validation is essential before any substantive conclusion or operational use. With verified data and appropriate governance, multi-source predictive analytics can support earlier and more targeted assistance while preserving educational judgment, fairness, and student agency.

ACKNOWLEDGMENT (IF ANY)

The authors acknowledge the researchers and higher-education practitioners whose work in educational data mining, learning analytics, student engagement, and explainable machine learning informed the development of this manuscript. Because the current version is a manuscript-development draft based on synthetic demonstration data, no real participants or institutions are represented in the reported results. Funding information, ethics approval, data-provider acknowledgments, and institutional support should be added only after an empirical study has been completed and verified.

AUTHOR CONTRIBUTIONS STATEMENT

AA: conceptualization, literature synthesis, research design, feature engineering framework, machine-learning analysis plan, visualization, and writing—original draft. BB: methodology review, validation strategy, interpretation, ethics and governance review, writing—review and editing, and supervision. Both authors approved the final manuscript and agree to be accountable for the integrity of the work. Replace AA and BB with the authors' actual initials before submission.

REFERENCES

- Alalawi, K., Athauda, R., & Chiong, R. (2023). Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review. *Engineering Reports*, 5(12), e12699. <https://doi.org/10.1002/eng2.12699>
- Alalawi, K., Athauda, R., Chiong, R., & Renner, I. (2025). Evaluating the student performance prediction and action framework through a learning analytics intervention study. *Education and Information Technologies*, 30, 2887–2916. <https://doi.org/10.1007/s10639-024-12923-5>
- Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student' performance prediction using machine learning techniques. *Education Sciences*, 11(9), 552. <https://doi.org/10.3390/educsci11090552>

- Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17, 3. <https://doi.org/10.1186/s41239-020-0177-7>
- Bergdahl, N., Bond, M., Sjöberg, J., Dougherty, M., & Oxley, E. (2024). Unpacking student engagement in higher education learning analytics: A systematic review. *International Journal of Educational Technology in Higher Education*, 21, 63. <https://doi.org/10.1186/s41239-024-00493-y>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, J., Zhou, X., Yao, J., & Tang, S.-K. (2025). Application of machine learning in higher education to predict students' performance, learning engagement and self-efficacy: A systematic literature review. *Asian Education and Development Studies*, 14(2), 205–240. <https://doi.org/10.1108/AEDS-08-2024-0166>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297. <https://doi.org/10.1007/BF00994018>
- Goh, T.-T. (2026). Learning management system log analytics: The role of persistence and consistency of engagement behaviour on academic success. *Journal of Computers in Education*, 13, 283–306. <https://doi.org/10.1007/s40692-025-00358-x>
- Hernández-García, Á., Cuenca-Enrique, C., Del-Río-Carazo, L., & Iglesias-Pradas, S. (2024). Exploring the relationship between LMS interactions and academic performance: A Learning Cycle approach. *Computers in Human Behavior*, 155, 108183. <https://doi.org/10.1016/j.chb.2024.108183>
- Herrmann, L., & Weigert, J. (2024). AI-based prediction of academic success: Support for many, disadvantage for some? *Computers and Education: Artificial Intelligence*, 7, 100303. <https://doi.org/10.1016/j.caeai.2024.100303>
- Johora, F. T., Hasan, M. N., Rajbongshi, A., Ashrafuzzaman, M., & Akter, F. (2025). An explainable AI-based approach for predicting undergraduate students academic performance. *Array*, 26, 100384. <https://doi.org/10.1016/j.array.2025.100384>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Wang, S., & Luo, B. (2024). Academic achievement prediction in higher education through interpretable modeling. *PLOS ONE*, 19(9), e0309838. <https://doi.org/10.1371/journal.pone.0309838>
- York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research & Evaluation*, 20(5), 1–20.