



PREDICTING STUDENT ACADEMIC RISK IN MOODLE LMS: A COMPARATIVE MACHINE LEARNING APPROACH TO EDUCATIONAL DATA MINING

Egamberdiyeva Shaxina Ulugbek qizi - Master's student at Bukhara State
University

Abstract. While Moodle-integrated monitoring dashboards can present instructors with descriptive indicators of student performance, they remain fundamentally retrospective: they report what has already happened rather than forecasting what is about to happen. This study addresses that limitation by developing and comparing five supervised machine learning classifiers — Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Extreme Gradient Boosting (XGBoost) — for the early prediction of academic risk from Moodle log and gradebook data. A dataset of 1,248 student-course records spanning eight courses over three academic semesters at a technical university in Uzbekistan was extracted via the Moodle Web Services API and engineered into 22 behavioral and performance features, including login regularity, resource-access latency, quiz-attempt patterns, and assignment-submission timeliness. Models were trained and validated using stratified 10-fold cross-validation and evaluated with accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). XGBoost achieved the strongest overall performance (accuracy = 90.6%, F1 = 0.887, AUC = 0.934), outperforming Random Forest (AUC = 0.921) and substantially outperforming Logistic Regression (AUC = 0.842). Critically, when restricted to data available during only the first four weeks of the semester, the ensemble models still achieved an AUC above 0.85, indicating that meaningful risk prediction is possible well before traditional grading milestones. SHAP-based feature-importance analysis identified early quiz performance, resource-access regularity, and assignment-submission delay as the three most influential predictors across models.



These findings demonstrate that ensemble tree-based methods applied to routinely collected Moodle log data can provide a computationally inexpensive yet accurate early-warning mechanism, complementing rule-based monitoring tools with genuine predictive capability.

Keywords: Moodle LMS, educational data mining, machine learning, academic risk prediction, XGBoost, random forest, feature importance, learning analytics.

1. Introduction

Early identification of students at risk of academic failure is one of the most consistently cited goals of learning analytics research, yet most operational monitoring tools deployed within Learning Management Systems remain descriptive rather than predictive. Threshold-based dashboards — including composite indices built from weighted averages of quiz, assignment, and engagement data — are transparent and easy for instructors to interpret, but they treat all indicators as linearly and independently related to outcome, an assumption rarely satisfied by real behavioral data, and they typically require a substantial accumulation of grades before a reliable signal emerges.

Machine learning offers an alternative: rather than defining risk through a fixed formula, a classifier can be trained to learn the (often nonlinear and interactive) relationship between behavioral and performance features recorded early in a course and the eventual academic outcome. This approach has been explored extensively in the international educational data mining literature, with tree-ensemble methods such as Random Forest and Gradient Boosting repeatedly reported as strong performers for tabular, feature-engineered LMS data. However, evidence specific to Moodle deployments in Central Asian higher-education contexts — where course structures, assessment cadences, and student engagement patterns may differ from the North American and Western European settings that dominate the published literature — remains limited.



This study addresses that gap by asking three research questions. RQ1: Among commonly used supervised classifiers, which achieves the highest predictive accuracy for end-of-semester academic risk using features derived from Moodle log and gradebook data? RQ2: How early in the semester can acceptable predictive performance be achieved using only data available up to that point? RQ3: Which behavioral and performance features contribute most strongly to model predictions, and are these consistent with the pedagogical intuitions underlying simpler rule-based monitoring tools? The remainder of the paper proceeds through the standard IMRAD structure: Section 2 details the dataset, feature engineering, modeling pipeline, and evaluation protocol; Section 3 reports comparative model performance, early-prediction results, and feature-importance findings; Section 4 discusses these results in relation to prior work and their practical implications; and Section 5 concludes.

2. Materials and Methods

2.1. Dataset and Feature Engineering

Data were extracted through the Moodle Web Services REST API from eight courses (computer engineering, applied mathematics, and general education) over three consecutive academic semesters at a technical university in Uzbekistan, yielding 1,248 anonymized student-course records after removing incomplete cases (7.1% of the original 1,343 records, mainly students who withdrew before mid-semester). The binary outcome label — Academic Risk (1) versus No Risk (0) — was assigned retrospectively based on final course grade, with Risk defined as a final grade below the institutional passing threshold or a grade in the lowest performance decile for pass/fail courses; the resulting classes were moderately imbalanced (22.4% Risk, 77.6% No Risk).

Twenty-two features were engineered from raw log and gradebook tables, grouped into four categories: (a) engagement features — weekly login count, session-length variability, and resource-view frequency relative to course peers; (b) assessment features — mean and variance of quiz scores, number of quiz attempts,



and assignment grades to date; (c) timeliness features — mean and maximum delay (in hours) between assignment deadline and actual submission, and proportion of on-time submissions; and (d) participation features — forum post count and reply latency. All continuous features were standardized (zero mean, unit variance) prior to model training; categorical course identifiers were one-hot encoded. Class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE) applied only within training folds to avoid data leakage.

2.2. Classification Algorithms

Five supervised classifiers commonly reported in the educational data mining literature were implemented and compared: (1) Logistic Regression, as an interpretable linear baseline; (2) a single Decision Tree (CART, Gini impurity criterion); (3) Random Forest, an ensemble of 300 bagged decision trees with feature subsampling; (4) Support Vector Machine with a radial basis function (RBF) kernel; and (5) Extreme Gradient Boosting (XGBoost), a regularized gradient-boosted tree ensemble. Hyperparameters for each model (e.g., tree depth, number of estimators, regularization strength, SVM kernel coefficient) were tuned via grid search with nested 5-fold cross-validation on the training partition to avoid optimistic bias in the reported test performance.

2.3. Evaluation Protocol

Model performance was assessed using stratified 10-fold cross-validation on the full dataset, with accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) reported as the mean across folds. Because early identification is the primary practical use case, a second analysis restricted the feature set to only those values observable within the first four weeks of a 15-week semester (login and engagement data, first quiz attempt, and first assignment submission) and re-evaluated all five models under the same cross-validation protocol to assess early-prediction feasibility (RQ2). Finally, SHAP (SHapley Additive exPlanations) values were computed for the best-performing



model to quantify each feature's marginal contribution to individual predictions and to rank global feature importance (RQ3).

3. Results

3.1. Comparative Classifier Performance

Table 1 reports full-semester performance for all five classifiers. XGBoost achieved the highest scores on every metric (accuracy = 90.6%, precision = 0.869, recall = 0.906, F1 = 0.887, AUC = 0.934), followed closely by Random Forest (AUC = 0.921). Both tree-ensemble methods outperformed the single Decision Tree (AUC = 0.849) by a wide margin, confirming the expected benefit of ensembling for this feature set. The RBF-kernel SVM (AUC = 0.878) outperformed the Logistic Regression baseline (AUC = 0.842), suggesting genuine nonlinearity in the relationship between behavioral features and risk that linear models fail to capture fully. Differences between XGBoost and Random Forest were not statistically significant (paired t-test on fold-wise AUC, $p = 0.19$), whereas both significantly outperformed Logistic Regression and the single Decision Tree ($p < 0.01$).

Model	Accuracy	Precision	Recall	AUC-ROC
Logistic Regression	79.3%	0.741	0.702	0.842
Decision Tree	81.0%	0.756	0.769	0.849
Support Vector Machine (RBF)	84.7%	0.803	0.811	0.878
Random Forest	89.1%	0.852	0.887	0.921
XGBoost	90.6%	0.869	0.906	0.934

Table 1. Mean 10-fold cross-validated performance of five classifiers using full-semester features.

3.2. Early-Prediction Feasibility

When restricted to features available within the first four weeks of the semester, predictive performance declined for all models but remained practically



useful for the ensemble methods: XGBoost retained an AUC of 0.861 and Random Forest an AUC of 0.853, compared with 0.934 and 0.921 respectively using full-semester data. Logistic Regression and the single Decision Tree degraded more sharply (AUC 0.774 and 0.781, respectively), indicating that the ensemble models were better able to extract signal from the sparser, noisier early-semester feature set. These results support a positive answer to RQ2: acceptable early prediction (AUC > 0.85) is achievable using ensemble methods roughly ten weeks before the traditional end-of-semester grading point.

3.3. Feature Importance

SHAP analysis of the XGBoost model identified first-quiz score, resource-access regularity (coefficient of variation of weekly login count), and mean assignment-submission delay as the three highest-magnitude predictors, jointly accounting for approximately 54% of total SHAP value magnitude across the test set. Forum participation features contributed comparatively little (combined SHAP contribution below 6%), suggesting that, in this dataset, passive engagement and assessment timeliness were considerably more diagnostic of eventual risk than active discussion-forum participation — a finding with direct implications for which indicators simpler rule-based monitoring tools should prioritize.

4. Discussion

The comparative results confirm, in a Moodle-based Central Asian university setting, a pattern well established in the broader educational data mining literature: tree-ensemble methods such as Random Forest and gradient-boosted trees consistently outperform linear and single-tree baselines on tabular LMS-derived features, and the performance gap is large enough to be pedagogically consequential rather than merely statistically significant. The finding that XGBoost sustained an AUC above 0.85 using only the first four weeks of data (RQ2) is particularly relevant for practice, since it indicates that predictive — rather than purely retrospective — early-warning systems are feasible using data every Moodle installation already collects by default, without requiring additional instrumentation.



The feature-importance results (RQ3) provide a partial validation, and a partial correction, of the assumptions embedded in simpler composite-index monitoring tools: early quiz performance and submission timeliness were confirmed as strong predictors, consistent with their prominent weighting in rule-based indices, but login-regularity emerged as more informative than raw activity-count or forum-participation measures, which are more commonly emphasized in engagement-scoring formulas. This suggests that institutions relying on hand-weighted composite indices could improve predictive validity by re-weighting toward access regularity rather than volume of activity.

Limitations include the single-institution, three-semester scope of the dataset, which may limit generalizability to institutions with different course structures or assessment cultures; the moderate class imbalance, which was addressed statistically via SMOTE but could still bias precision estimates for the minority (at-risk) class; and the retrospective nature of the outcome label, which does not capture students who received informal, undocumented support during the semester and whose final grade may therefore understate their true initial risk trajectory. In addition, while SHAP values support local and global interpretability, deploying a black-box ensemble model in a live advising workflow raises fairness and accountability considerations — particularly the risk that model errors could disproportionately affect specific student subgroups — that were outside the scope of this study but warrant dedicated investigation, including subgroup performance audits, before operational deployment.

Future work should validate these models on multi-institutional data, evaluate fairness metrics across demographic and prior-achievement subgroups, and integrate the trained classifier as a predictive layer within the rule-based monitoring dashboard described in related work, allowing instructors to view both an interpretable composite index and a calibrated probabilistic risk score side by side.



5. Conclusion

This study compared five supervised machine learning classifiers for predicting student academic risk from Moodle log and gradebook data across 1,248 student-course records. Extreme Gradient Boosting achieved the strongest performance (AUC = 0.934 using full-semester data; AUC = 0.861 using only the first four weeks), significantly outperforming linear and single-tree baselines, with first-quiz performance, login regularity, and assignment-submission timeliness identified as the most influential predictors. These findings demonstrate that accurate, genuinely predictive — rather than purely descriptive — academic early-warning capability can be built from data that Moodle already collects by default, offering universities a computationally inexpensive complement to rule-based monitoring dashboards and a concrete direction for strengthening evidence-based academic advising in resource-constrained higher-education settings such as Uzbekistan.

REFERENCES

1. 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.
2. Marbouti, F., Diefes-Dux, H. A., Madhavan, K. (2016). Models for Early Prediction of At-Risk Students in a Course Using Standards-Based Grading. *Computers & Education*, 103, 1–15.
3. Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., Rego, J. (2017). Evaluating the Effectiveness of Educational Data Mining Techniques for Early Prediction of Students' Academic Failure in Introductory Programming Courses. *Computers in Human Behavior*, 73, 247–256.
4. Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., Nawaz, R. (2020). Predicting Academic Performance of Students from VLE Big Data Using Deep Learning Models. *Computers in Human Behavior*, 104, 106189.



5. 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, 6347186.
6. Xing, W., Du, D. (2019). Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. *Journal of Educational Computing Research*, 57(3), 547–570.
7. Romero, C., Ventura, S. (2020). Educational Data Mining and Learning Analytics: An Updated Survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), e1355.
8. Baker, R. S., Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In *Learning Analytics: From Research to Practice* (pp. 61–75). Springer.
9. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
10. 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.
11. Lundberg, S. M., Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
12. Moodle Pty Ltd. (2024). Moodle Web Services API Documentation. Retrieved from https://docs.moodle.org/dev/Web_service_API_functions
13. Karimov, I., Yusupov, R. (2022). Digitalization of Higher Education in Uzbekistan: Current State and Prospects. *Journal of Educational Technology Development and Exchange*, 15(2), 45–58.