Journal of Electrical Engineering and Computer (JEECOM)
Vol. 8, No. 1 (2026), DOI: 10.33650/jeecom.v8i1.16978
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Explainable Machine Learning for Predicting Student Dropout and Academic Success Using XGBoost and SHAP

Sidik Praptomo [1*], Ahmad Risman [2], Riko Muhammad Suri [3]
[1,2,3] Universitas Muhammadiyah Muara Bungo, Jambi, Indonesia

Article Info

Article history:

Received April 2, 2026
Revised April 26, 2026
Accepted April 28, 2026

Keywords:

Educational Data Mining
Student Dropout
XGBoost
Explainable Artificial Intelligence
SHAP

ABSTRACT

Student dropout is a persistent challenge in higher education, and predictive models can support early identification of students who may require academic or financial intervention. This study develops an explainable multiclass machine learning approach to predict three academic outcomes—Dropout, Enrolled, and Graduate—using the public Predict Students' Dropout and Academic Success dataset containing 4,424 student records. Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost) were compared using a stratified 80:20 hold-out design. XGBoost hyperparameters were optimized through randomized search with five-fold stratified cross-validation, and SHapley Additive exPlanations (SHAP) were used to interpret global and class-specific predictions. Random Forest achieved the highest overall accuracy of 77.18%, whereas the optimized XGBoost model produced the highest macro recall of 69.58% and macro F1-score of 70.08%. XGBoost improved recall for the minority Enrolled class to 46.54%, compared with 38.36% for Random Forest and 33.33% for Logistic Regression. SHAP analysis showed that curricular units approved in the second and first semesters, tuition-fee status, course, second-semester grade, and age at enrollment had the largest contributions to the model outputs. Low academic progression and unpaid tuition status were associated with positive contributions toward Dropout predictions, while stronger academic progression contributed toward Graduate predictions. These findings show that explainability complements predictive evaluation by revealing potentially useful patterns for educational decision support.

Corresponding Author:
Sidik Praptomo,
Universitas Muhammadiyah Muara bungo, Jalan Pangeran Diponegoro, 24 37218, Muara Bungo, Jambi
Email: sidikpraptomo7@gmail.com; ORCID iD: http://orcid.org/0009-0002-2875-400X

1. INTRODUCTION

Student dropout is a multidimensional problem in higher education because it affects students, institutions, and the efficiency of educational resources. Early identification of students at risk can support targeted academic advising, financial assistance, and retention strategies before academic difficulties become irreversible. Previous studies have therefore increasingly treated dropout prediction as an educational data mining task in which administrative, demographic, socioeconomic, and academic variables are used to estimate students' future academic outcomes [1]–[3].

Machine learning has broadened the range of methods available for student-outcome prediction. Tree-based ensembles, boosting algorithms, neural models, and stacked approaches have been evaluated across different higher-education contexts [4]–[7]. However, strong overall performance does not by itself resolve two issues that are important in a three-class setting: whether recognition is distributed evenly across Dropout, Enrolled, and Graduate students, and whether the features used by a model contribute in the same way to each

predicted outcome. These issues are especially relevant when one class is substantially smaller than the others and aggregate accuracy can conceal weak class-level performance [2], [5].

A widely used benchmark for this problem is the Predict Students' Dropout and Academic Success dataset introduced by Realinho et al. [1]. The dataset contains demographic, socioeconomic, enrollment, and academic information from 4,424 students and defines three target outcomes: Dropout, Enrolled, and Graduate. Martins et al. later examined phased multiclass prediction using the same institutional context and demonstrated that academic information collected during the first semesters provides substantial predictive value [2]. The dataset is particularly useful for reproducible educational data mining research because it allows different models to be evaluated on a common multiclass task.

However, predictive accuracy alone is insufficient when a model is intended to support educational decision-making. Complex ensemble models may provide strong predictive capability while offering limited transparency about why a student is classified into a particular outcome. Explainable Artificial Intelligence (XAI) addresses this limitation by providing methods that expose the contribution of individual predictors to model outputs [8]–[11]. In educational applications, interpretability is important because institutional interventions should be based on understandable evidence rather than opaque risk scores.

SHapley Additive exPlanations (SHAP) provide a theoretically grounded method for explaining model predictions by assigning contribution values to input features. SHAP can be used at a global level to identify features that generally influence a model and at a class or instance level to examine the direction and magnitude of feature effects. Such explanations are particularly suitable for tree-based models such as XGBoost, which can capture nonlinear interactions but are less directly interpretable than conventional statistical models [12], [13].

Recent dropout-prediction research has increasingly combined predictive modeling with explainability. Nagy and Molontay examined interpretable dropout prediction for personalized intervention [8], Zanellati et al. studied the trade-off between predictive performance and explainability [9], Villar and de Andrade compared a broad set of supervised algorithms and incorporated SHAP into student-outcome analysis [7], and more recent work has combined XGBoost with explainability for early-warning applications [10]. These studies establish that both predictive performance and interpretability are important. What remains less emphasized, particularly for the three-class public benchmark used in this study, is their joint examination under the original class proportions: how the models behave for the underrepresented Enrolled class and whether the explanatory patterns differ across Dropout, Enrolled, and Graduate predictions rather than being discussed only at the aggregate model level.

This study does not propose a new classification algorithm. Its contribution is analytical: it preserves the original class distribution, compares Logistic Regression, Random Forest, and XGBoost using both overall and macro-averaged metrics, and then interprets the optimized XGBoost model at global and class-specific levels using SHAP. Particular attention is given to the Enrolled class because it is the smallest class in the dataset and the most difficult to recognize in the present experiments. By linking class-level performance with class-specific explanations, the study aims to clarify a trade-off that would be hidden if model selection were based on accuracy alone.

Accordingly, this study addresses three research questions: (RQ1) how do Logistic Regression, Random Forest, and XGBoost compare in predicting Dropout, Enrolled, and Graduate outcomes under the original class distribution; (RQ2) how does recognition differ across the three classes, particularly for the underrepresented Enrolled class; and (RQ3) which academic, financial, demographic, and enrollment-related features have the largest SHAP contributions to each predicted outcome?

2. METHOD

2.1. Dataset and Target Classes

The experiment used the public Predict Students' Dropout and Academic Success dataset created for early identification of academic dropout and failure in higher education [1]. The experiment retrieved the original UCI repository version through the ucimlrepo interface in Google Colab. The dataset contains 4,424 records and 36 predictor variables representing demographic characteristics, socioeconomic conditions, enrollment information, and academic performance. The target variable contains three classes: Graduate, Dropout, and Enrolled. The class distribution was preserved without oversampling or undersampling so that the evaluation reflected the original benchmark distribution.

Table 1. Distribution of the target classes

Class	Number of Students	Percentage
Graduate	2,209	49.93%

ISSN: 2715-6427

Dropout	1,421	32.12%
Enrolled	794	17.95%
Total	4,424	100.00%

2.2. Preprocessing and Experimental Design

All experiments were conducted in Google Colab using Python with scikit-learn, XGBoost, and SHAP. The target labels were converted into machine-readable class codes using a label encoder. The dataset was divided into 80% training data and 20% testing data using a stratified hold-out split with random_state = 42, preserving the class proportions in both partitions. The resulting test set contained 885 students: 284 Dropout, 159 Enrolled, and 442 Graduate. Median imputation was included in the pipelines as a safeguard for missing numeric values. Standard scaling was applied to Logistic Regression, whereas tree-based models were trained without feature scaling. The original numeric coding of categorical variables was retained to remain consistent with the public benchmark representation; therefore, nominal variables such as Course are interpreted as categories rather than as ordinal numeric scales.

2.3. Classification Models

Three supervised classifiers were evaluated. Logistic Regression was used as a linear baseline because of its simplicity and direct probabilistic classification capability. Random Forest was used as a bagging-based ensemble comparator with 400 trees and balanced class weights. XGBoost was selected as the primary model because gradient-boosted decision trees are effective at modeling nonlinear relationships and interactions in structured tabular data [12]. Rather than assuming that the model with the highest accuracy would be the most useful, all three models were evaluated using macro-averaged metrics so that performance on the minority Enrolled class contributed equally to the overall assessment.

2.4. XGBoost Hyperparameter Optimization

XGBoost hyperparameters were optimized only on the training partition using RandomizedSearchCV. Fifteen candidate configurations were sampled and evaluated with stratified five-fold cross-validation, producing 75 model fits. Macro F1-score was used as the optimization objective because it assigns equal importance to all three target classes and is less dominated by the majority Graduate class than overall accuracy. The best cross-validation Macro F1-score was 0.7144. The final optimized configuration is shown in Table 2.

Table 2. Best XGBoost hyperparameters from randomized search

Hyperparameter	Best Value
n_estimators	500
max_depth	5
learning_rate	0.1
min_child_weight	3
subsample	1.0
colsample_bytree	0.9
gamma	0

2.5. Evaluation Metrics

Model performance was evaluated using accuracy, macro precision, macro recall, macro F1-score, class-specific precision, recall, and F1-score, together with confusion matrices. Accuracy measures the proportion of correctly classified test instances, while macro-averaged metrics calculate the unweighted mean across classes. Macro F1-score was treated as the primary balanced metric because the target distribution is unequal and the Enrolled class represents only 17.95% of the data. This choice prevents strong performance on the majority Graduate class from masking weak minority-class recognition, a known concern in imbalanced educational datasets [7], [14].

2.6. SHAP-Based Explainability

The optimized XGBoost model was interpreted using SHAP, which attributes each prediction to the contribution of individual features [13]. TreeExplainer was applied to a reproducible random sample of 500 observations from the test set. Global importance was calculated from the mean absolute SHAP values across

observations and classes. Class-specific SHAP summary plots were then analyzed for Dropout, Enrolled, and Graduate to identify both the magnitude and direction of influential features. SHAP values were interpreted as explanations of model behavior rather than evidence of causal relationships.

3. RESULTS AND DISCUSSION

3.1. Target-Class Distribution

The dataset exhibits a moderate multiclass imbalance. Graduate students account for 49.93% of all observations, followed by Dropout at 32.12% and Enrolled at only 17.95%. Figure 1 visualizes this distribution. The relatively small Enrolled class is important when interpreting accuracy because a model can achieve a favorable overall score while still performing poorly on students whose academic status remains unresolved. This distribution motivated the use of macro-averaged metrics in addition to accuracy.

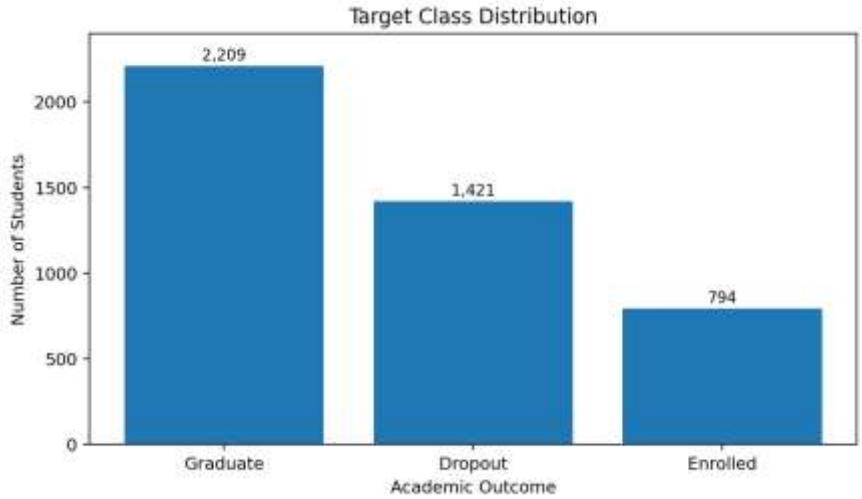

Figure 1. Distribution of the target classes

3.2. Comparative Model Performance

Table 3 presents the hold-out test performance of the three models. Random Forest achieved the highest overall accuracy at 0.7718 and the highest macro precision at 0.7275. In contrast, XGBoost achieved the highest macro recall (0.6958) and macro F1-score (0.7008). The Macro F1 advantage of XGBoost over Random Forest is small in absolute terms (0.0029), so the results do not support a claim that XGBoost dominates all metrics. Its main advantage is instead visible in the distribution of class-level recall, particularly for Enrolled students.

Table 3. Comparison of model performance on the hold-out test set

Model	Accuracy	Macro Precision	Macro Recall	Macro F1
XGBoost	0.7605	0.7075	0.6958	0.7008
Random Forest	0.7718	0.7275	0.6859	0.6979
Logistic Regression	0.7684	0.7070	0.6754	0.6826

ISSN: 2715-6427

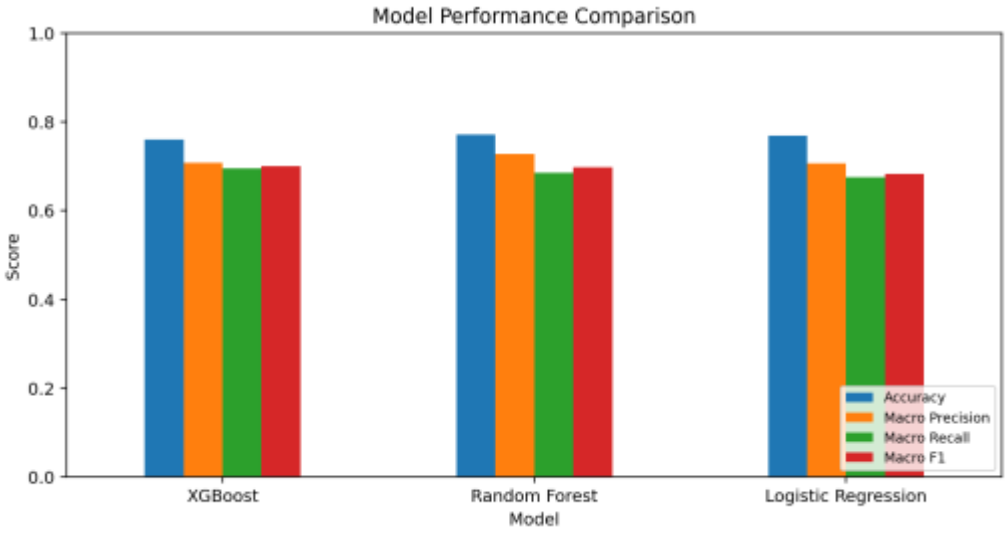

Figure 2. Comparison of accuracy and macro-averaged metrics

The result illustrates why accuracy should not be used in isolation for this dataset. Random Forest correctly classified the largest proportion of all students, but XGBoost distributed its predictive performance more evenly across the three outcomes. This observation is consistent with prior comparative studies showing that ensemble and boosting models can provide competitive results on student-outcome data, while the preferred model depends on the evaluation objective and treatment of class imbalance [5]–[7].

3.3. XGBoost Class-Specific Performance

The optimized XGBoost model achieved an accuracy of 76.05% on 885 hold-out observations. Its cross-validation Macro F1-score was 71.44%, compared with 70.08% on the independent test set, a difference of 1.36 percentage points. This relatively small reduction suggests reasonably stable generalization under the selected split, although broader validation across institutions would be required before claiming external generalizability.

Table 4. Class-specific performance of the optimized XGBoost model

Class	Precision	Recall	F1-score	Support
Dropout	0.7940	0.7465	0.7695	284
Enrolled	0.5103	0.4654	0.4868	159
Graduate	0.8182	0.8756	0.8459	442
Macro average	0.7075	0.6958	0.7008	885
Weighted average	0.7551	0.7605	0.7569	885

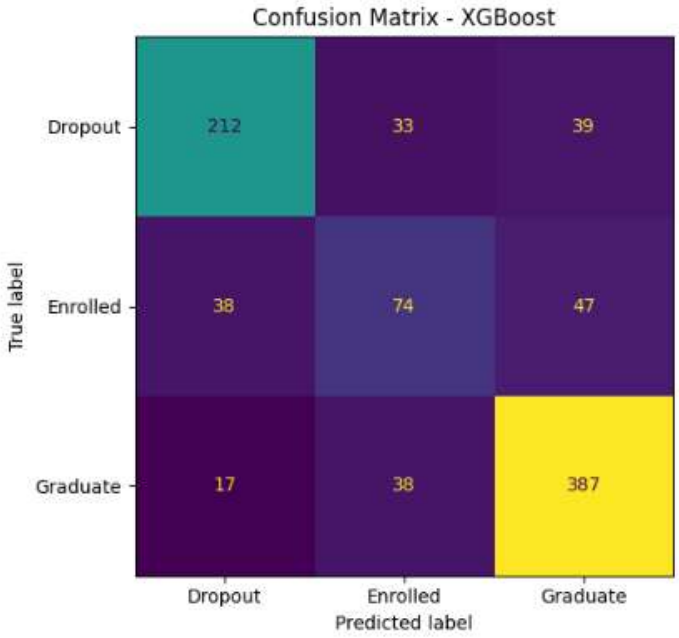

Figure 3. Confusion matrix of the optimized XGBoost model

Graduate was the easiest class for XGBoost to recognize, with recall of 87.56% and F1-score of 84.59%. Dropout was also predicted with relatively strong precision (79.40%), recall (74.65%), and F1-score (76.95%). The Enrolled class remained substantially more difficult, with recall of 46.54% and F1-score of 48.68%. From the 159 Enrolled students in the test set, 74 were correctly classified, 38 were predicted as Dropout, and 47 were predicted as Graduate. This distribution of errors shows that the model had greater difficulty separating Enrolled observations from the other two classes; it should not be interpreted as evidence that Enrolled forms a scientifically established intermediate state.

Despite this difficulty, XGBoost improved Enrolled recall compared with both baseline models: Logistic Regression achieved 33.33%, Random Forest achieved 38.36%, and XGBoost achieved 46.54%. Thus, XGBoost increased minority-class recall by 8.18 percentage points over Random Forest and 13.21 percentage points over Logistic Regression. This improvement explains why XGBoost obtained the highest macro recall and macro F1-score even though its overall accuracy was lower than Random Forest.

3.4. Global Feature Importance and SHAP Explanation

Global SHAP analysis provides a model-level description of the variables with the largest contributions to the XGBoost outputs across the three classes. As shown in Figure 4, Curricular units 2nd sem (approved) had the largest mean absolute SHAP contribution by a clear margin. It was followed by Curricular units 1st sem (approved), Tuition fees up to date, Course, Curricular units 2nd sem (grade), Age at enrollment, Admission grade, Curricular units 2nd sem (evaluations), Curricular units 2nd sem (enrolled), and Scholarship holder. This ranking indicates that the model relied more strongly on realized first-year academic progress than on most background characteristics, without implying a causal effect of these variables on student outcomes.

ISSN: 2715-6427

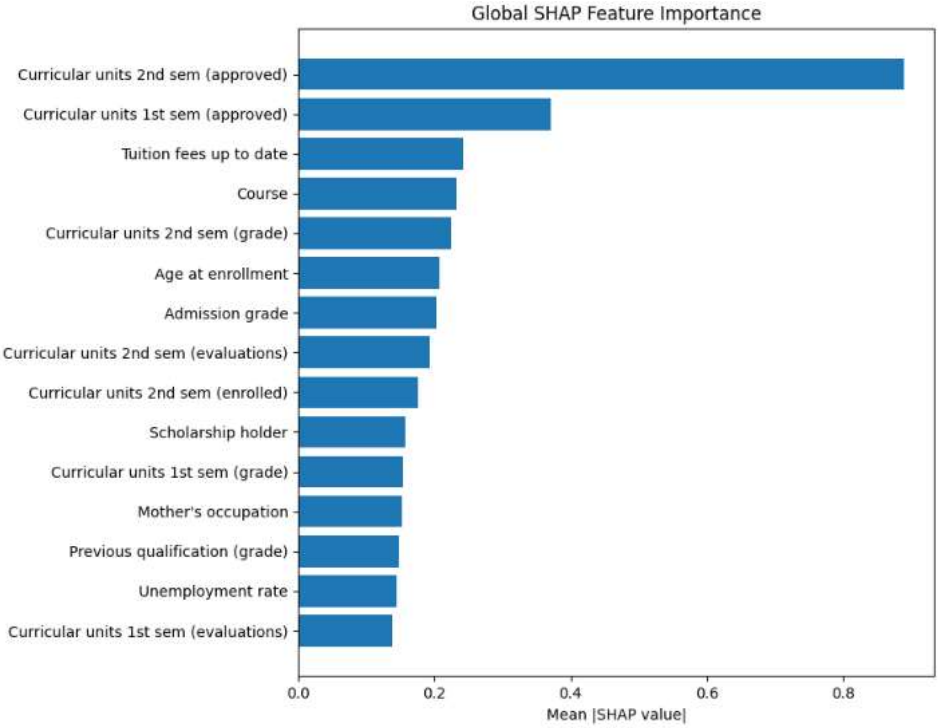

Figure 4. Global SHAP feature importance for the optimized XGBoost model

The importance of approved curricular units is consistent with prior research emphasizing academic achievement and progression as strong predictors of dropout and success [2], [3], [6]. Financial variables also remain relevant. Tuition fees up to date appears among the leading global predictors, suggesting that payment status carries information that complements academic performance. Course also ranks highly, but because it is a nominal code in the dataset, its numeric value must not be interpreted as an ordered scale; the result only indicates that differences among course categories influence model predictions.

3.5. Class-Specific SHAP Analysis

The Dropout SHAP summary plot in Figure 5 shows a strong directional pattern for academic progression. Low values of Curricular units 2nd sem (approved) are associated with positive SHAP values for the Dropout class, whereas higher values shift the prediction away from Dropout. A similar pattern is visible for first-semester approved units. Tuition fees up to date also produces a substantial separation: students whose fees are not current tend to receive positive contributions toward Dropout predictions. Higher age at enrollment shows a weaker but visible tendency toward positive Dropout contributions in this model. These patterns should be interpreted as associations learned by the model, not as causal effects.

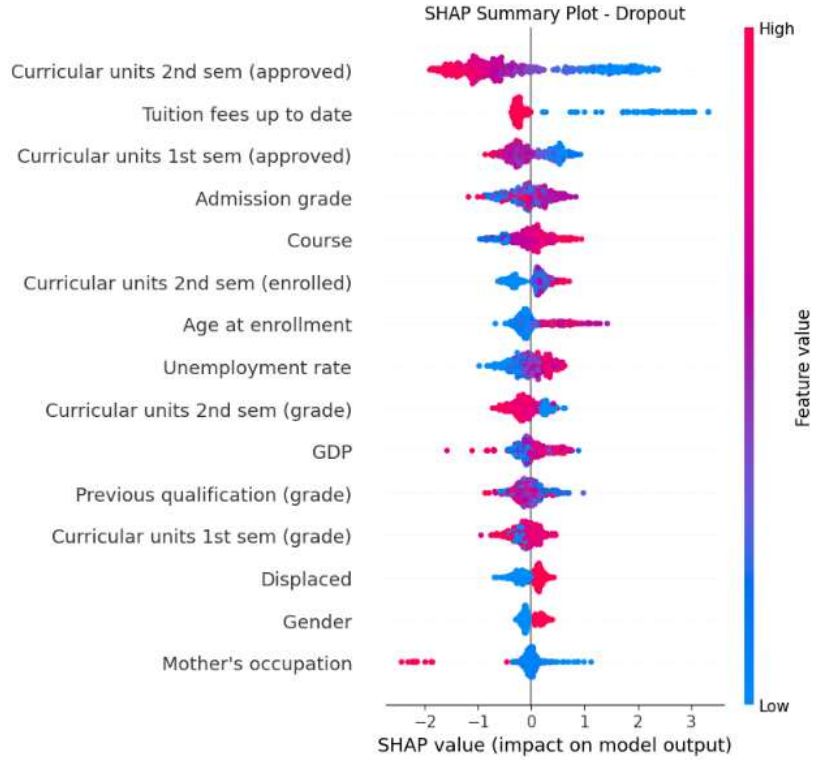

Figure 5. SHAP summary plot for the Dropout class

For the Enrolled class, the SHAP pattern is less polarized (Figure 6). The leading variables include second-semester evaluations, second-semester approved units, first-semester evaluations, second-semester grade, and first-semester approved units. Positive and negative SHAP values are intermingled across these academic indicators, while the confusion matrix shows that Enrolled observations were misclassified toward both Dropout and Graduate. Together, these results indicate weaker class separation for Enrolled in the fitted model. They do not establish that Enrolled is an intermediate academic state in a causal or developmental sense.

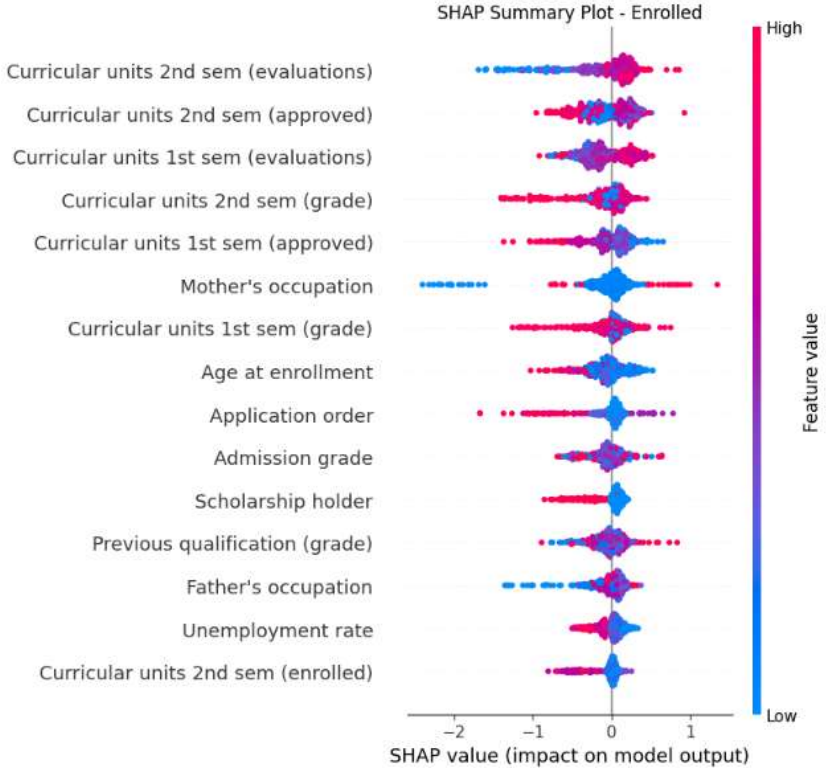

Figure 6. SHAP summary plot for the Enrolled class

ISSN: 2715-6427

The Graduate explanation in Figure 7 exhibits the reverse pattern of the Dropout class. High values of Curricular units 2nd sem (approved) and Curricular units 1st sem (approved) contribute positively toward Graduate predictions, while low values shift the prediction away from graduation. Scholarship holder status also tends to contribute positively, whereas unfavorable tuition-fee status and debtor status reduce the model's tendency to predict Graduate. The opposition between the Dropout and Graduate SHAP patterns increases confidence that the model is using academically coherent signals rather than relying solely on demographic proxies.

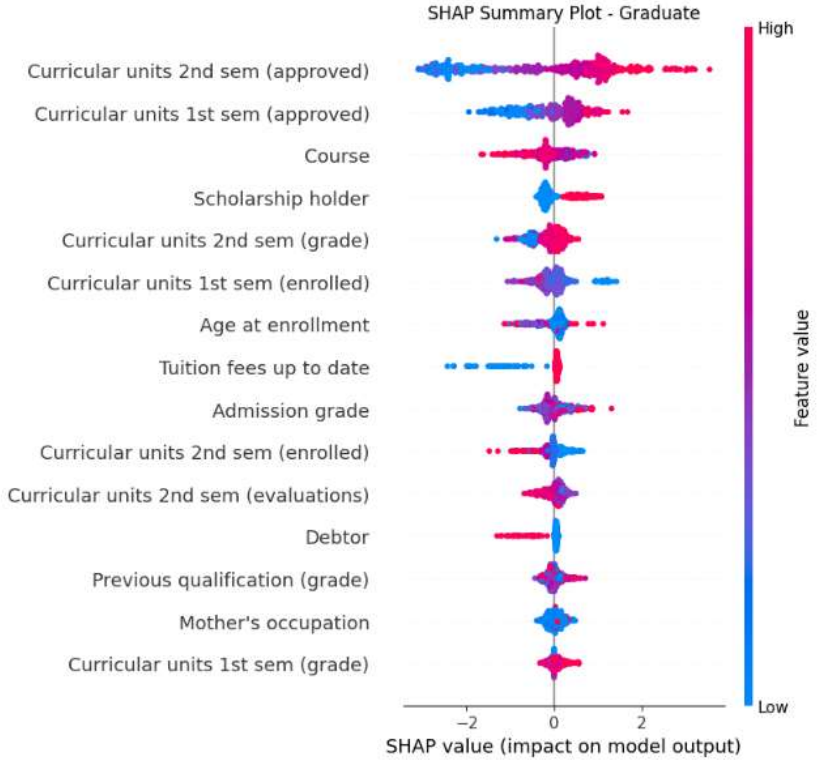

Figure 7. SHAP summary plot for the Graduate class

3.6. Discussion, Practical Implications, and Limitations

The predictive and SHAP results point to a consistent pattern in the fitted model. Academic progression variables, particularly approved curricular units, account for the largest contributions to both global and class-specific predictions. Tuition-fee status adds information that is distinct from grades and curricular progress, while the Enrolled class remains the least clearly separated outcome. In practice, these findings suggest that any decision-support use of the model should examine class-level performance together with the explanation attached to a prediction, rather than treating the overall accuracy as sufficient evidence for intervention.

The results are broadly consistent with the literature emphasizing boosting models, academic progression data, and interpretable machine learning for retention-oriented decision support [7]–[10]. The present analysis adds a narrower perspective by retaining the benchmark's original class proportions and explicitly examining the underrepresented Enrolled class alongside class-specific SHAP explanations. The resulting trade-off is clear: Random Forest provides slightly higher overall accuracy, whereas XGBoost shows only a small Macro F1 advantage but a more noticeable gain in Enrolled recall. Accordingly, the choice between models depends on whether the evaluation objective prioritizes overall correctness or more balanced recognition across outcome classes.

Several limitations should be considered. The dataset originates from a single Portuguese higher-education setting, so the learned relationships may not transfer directly to other institutions, countries, or academic systems. The public dataset uses integer codes for several nominal variables; although these codes are suitable for tree-based partitioning, directional SHAP interpretation is inappropriate for categorical variables such as Course and parental occupation. In addition, Random Forest used balanced class weights whereas XGBoost was optimized without resampling or class weighting, meaning that the comparison reflects

model-specific pipelines rather than identical imbalance treatments. Finally, the study uses one stratified hold-out split; future work should evaluate repeated or nested cross-validation, temporal validation, probability calibration, and institution-level external validation.

4. CONCLUSION

This study evaluated Logistic Regression, Random Forest, and XGBoost for multiclass prediction of Dropout, Enrolled, and Graduate outcomes and used SHAP to examine the optimized XGBoost model. Random Forest achieved the highest accuracy (77.18%), while XGBoost produced a slightly higher Macro F1-score (70.08% versus 69.79%) and, more importantly, a higher recall for the underrepresented Enrolled class (46.54% versus 38.36% for Random Forest and 33.33% for Logistic Regression). SHAP showed that approved curricular units in the first and second semesters had the largest contributions to model predictions, while tuition-fee status also contributed to Dropout-related outputs. These results indicate that model selection based only on accuracy can overlook class-level behavior; for educational decision support, balanced metrics and class-specific explanations should be considered together.

ACKNOWLEDGEMENTS

The authors acknowledge the creators of the Predict Students' Dropout and Academic Success dataset and the UCI Machine Learning Repository for providing open access to the data used in this study. The authors also acknowledge the open-source Python, scikit-learn, XGBoost, and SHAP communities that supported the reproducible experimental workflow.

REFERENCES

[1] V. Realinho, J. Machado, L. Baptista, and M. V. Martins, “Predicting Student Dropout and Academic Success,” Data, vol. 7, no. 11, Art. no. 146, 2022, doi: 10.3390/data7110146.
[2] M. V. Martins, L. Baptista, J. Machado, and V. Realinho, “Multi-Class Phased Prediction of Academic Performance and Dropout in Higher Education,” Applied Sciences, vol. 13, no. 8, Art. no. 4702, 2023, doi: 10.3390/app13084702.
[3] M. Cannistrà, C. Masci, F. Ieva, T. Agasisti, and A. M. Paganoni, “Early-predicting dropout of university students: an application of innovative multilevel machine learning and statistical techniques,” Studies in Higher Education, vol. 47, no. 9, pp. 1935–1956, 2022, doi: 10.1080/03075079.2021.2018415.
[4] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, “Predicting student’s dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization,” Computers and Education: Artificial Intelligence, vol. 3, Art. no. 100066, 2022, doi: 10.1016/j.caeai.2022.100066.
[5] Z. Song, S.-H. Sung, D.-M. Park, and B.-K. Park, “All-Year Dropout Prediction Modeling and Analysis for University Students,” Applied Sciences, vol. 13, no. 2, Art. no. 1143, 2023, doi: 10.3390/app13021143.
[6] M. Vaarma and H. Li, “Predicting student dropouts with machine learning: An empirical study in Finnish higher education,” Technology in Society, vol. 76, Art. no. 102474, 2024, doi: 10.1016/j.techsoc.2024.102474.
[7] A. Villar and C. R. V. de Andrade, “Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study,” Discover Artificial Intelligence, vol. 4, Art. no. 2, 2024, doi: 10.1007/s44163-023-00079-z.
[8] M. Nagy and R. Molontay, “Interpretable Dropout Prediction: Towards XAI-Based Personalized Intervention,” International Journal of Artificial Intelligence in Education, vol. 34, no. 2, pp. 274–300, 2024, doi: 10.1007/s40593-023-00331-8.
[9] A. Zanellati, S. P. Zingaro, and M. Gabbrielli, “Balancing Performance and Explainability in Academic Dropout Prediction,” IEEE Transactions on Learning Technologies, vol. 17, pp. 2086–2099, 2024, doi: 10.1109/TLT.2024.3425959.
[10] B. Carballo-Mendívil, A. Arellano-González, N. J. Ríos-Vázquez, and M. del P. Lizardi-Duarte, “Predicting Student Dropout from Day One: XGBoost-Based Early Warning System Using Pre-Enrollment Data,” Applied Sciences, vol. 15, no. 16, Art. no. 9202, 2025, doi: 10.3390/app15169202.
[11] H. Khosravi et al., “Explainable Artificial Intelligence in education,” Computers and Education: Artificial Intelligence, vol. 3, Art. no. 100074, 2022, doi: 10.1016/j.caeai.2022.100074.
[12] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[13] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 4765–4774.
[14] N. Mduma, “Data Balancing Techniques for Predicting Student Dropout Using Machine Learning,” Data, vol. 8, no. 3, Art. no. 49, 2023, doi: 10.3390/data8030049.