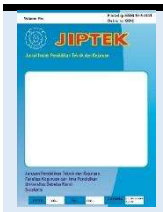

JIPTEK: Jurnal Ilmiah Pendidikan Teknik dan Kejuruan
Jurnal Homepage: https://jurnal.uns.ac.id/jptk

Explainable Machine Learning for Predicting Student Dropout Risk in Vocational Higher Education

M. Fajar Ramadhan[1*], Febriyanti Panjaitan[2], Winarnie[3], Hery Oktafiandi[4], Yohanes Yohanes[5]
[1*,2,3,4,5] Program Studi Teknik Informatika, Universitas Satu, Palembang, Indonesia
Email: m.ramadhan@univ.satu.ac.id, pebrianti.panjaitan@univ.satu.ac.id, winarnie@univ.satu.ac.id, hery.oktafiandy@univ.satu.ac.id, yohanes@univ.satu.ac.id

ABSTRACT

Student attrition remains a major concern in vocational higher education because it may disrupt study completion, academic support planning, and institutional quality assurance. This study develops an explainable predictive model to identify students with a higher probability of dropout. The analysis used the Predict Students' Dropout and Academic Success dataset from the UCI Machine Learning Repository, which contains 4,424 student records covering academic, demographic, socioeconomic, and early academic performance variables. The original outcome variable was converted from three categories into a binary target consisting of Dropout and Non-dropout. Logistic Regression, Random Forest, and XGBoost were trained and compared using accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, confusion matrix, and McNemar's test. To support interpretability, permutation feature importance and SHAP were applied to identify the main contributors to model predictions. The findings indicate that XGBoost produced the strongest descriptive performance, with an accuracy of 0.8859, an F1-score of 0.8250 for the Dropout class, and an ROC-AUC of 0.9355. Nevertheless, McNemar's test showed that its difference from Logistic Regression was not statistically significant. The most influential predictors were approved curricular units in the second semester, tuition fee payment status, and approved curricular units in the first semester.

Keywords: *student dropout; explainable machine learning; XGBoost; SHAP; vocational higher education*

JIPTEK: Jurnal Ilmiah Pendidikan Teknik dan Kejuruan
Vol 19 Issue 2 2026
DOI: https://doi.org/10.20961/jiptek.vyyix.xxxxx

INTRODUCTION

Vocational higher education plays a strategic role in preparing human resources with applied skills, work competence, and adaptive readiness for industrial needs. However, the success of higher education is not only measured by the admission of new students but also by the institution's ability to retain students until they complete their studies. Dropout is an important issue because it can generate academic, social, economic, and institutional consequences. Realinho et al. (2022, p. 1) explain that dropout and failure in higher education can hinder economic growth, competitiveness, and productivity, while also directly affecting

students, families, higher education institutions, and society.

In the context of higher education, dropout does not have a single universal definition. Realinho et al. (2022, p. 1) emphasize that there is no universal definition of dropout because its proportion may differ depending on data sources, measurement periods, and calculation methods. In vocational education, this issue also needs to be interpreted carefully because early termination does not always mean that students permanently leave the education system. Wydra-Somaggio (2021, pp. 1–2) shows that in vocational education, some learners may experience stopout, which means leaving one educational pathway but later re-entering vocational education in another field.

Student dropout is a multidimensional issue. Its causes are not only related to academic ability but also to financial, personal, family, socioeconomic, and institutional support factors. Martins et al. (2023, p. 1) explain that academic difficulties, financial problems, personal issues, and family conditions can affect students' academic performance and contribute to dropout. Rahmani et al. (2024, p. 1) also show that dropout in online higher education is influenced by demographic aspects, course characteristics, technology, motivation, support, health, financial constraints, isolation, and academic workload.

The development of learning analytics and educational data mining provides opportunities for educational institutions to use academic data as a basis for early detection. Yağcı (2022, p. 1) explains that educational data mining can be used to discover hidden relationships in educational data and predict students' academic achievement. De Oliveira et al. (2021, p. 1) also show that learning analytics can be used to analyze student retention and dropout through the use of educational data. At the end of their review, De Oliveira et al. (2021, p. 24) emphasize that learning analytics can help institutions identify students who may potentially drop out through risk indicators, although the use of personal data must still consider privacy issues.

Machine learning is widely used in dropout prediction because it can process various types of data, including academic, demographic, socioeconomic, and learning management system activity data. Realinho et al. (2022, pp. 2–3) explain that dropout prediction datasets may contain demographic, socioeconomic, macroeconomic, enrollment academic data, and academic performance data at the end of the first and second semesters. Vaarma and Li (2024, p. 1) also show that transcript, demographic, and LMS data can be used to predict student dropout in higher education. Thus, academic and learning behavior data can be used as a basis for developing early warning systems.

Several studies have compared machine learning algorithms for predicting dropout and academic success. Andrade-Girón et al. (2023, p. 1) show that Random Forest is one of the most commonly used algorithms in dropout prediction studies. Martins et al. (2023, p. 3) explain that several algorithms commonly used in predicting student performance and dropout include

Decision Tree, Support Vector Machine, Naive Bayes, Gradient Boosting, k-Nearest Neighbors, Random Forest, Artificial Neural Networks, and Deep Learning. Villar and de Andrade (2024, pp. 1–2) also compare various supervised learning algorithms, including Decision Tree, SVM, Random Forest, XGBoost, CatBoost, and LightGBM, to predict dropout and academic success.

Class imbalance is an important concern in dropout prediction. In many educational datasets, the number of students who drop out is smaller than the number of students who persist or graduate. Martins et al. (2023, pp. 2–3) state that minority classes are often the most relevant classes because students in these groups are those who most need pedagogical intervention. Villar and de Andrade (2024, pp. 2–3) also explain that class imbalance can make models biased toward the majority class and cause poor prediction of the minority class. Therefore, model evaluation should not rely only on accuracy but also consider precision, recall, F1-score, ROC-AUC, and PR-AUC.

XGBoost is used in this study because it is a tree-based boosting model widely used for tabular data and classification tasks. Chen and Guestrin (2016, p. 785) explain that XGBoost is a scalable tree boosting system widely used to achieve strong predictive results in various machine learning challenges. Villar and de Andrade (2024, pp. 1–2) also show that boosting models, such as XGBoost, CatBoost, and LightGBM, are relevant for dropout prediction and academic success, especially when studies deal with imbalanced data.

Although predictive performance is important, machine learning models in education should not stop at accuracy scores. Models need to be explainable so that the results can be understood by lecturers, tutors, study program managers, and policymakers. Nagy and Molontay (2024, pp. 274–277) show that the use of interpretable machine learning and explainable artificial intelligence in dropout prediction can help stakeholders understand the factors affecting student risk. Lundberg and Lee (2017, p. 1) explain that SHAP provides the contribution value of each feature to the model prediction. Nti and Ramanayake (2026, p. 1) also use SHAP to provide global and local interpretations so that educators can understand the factors driving dropout risk.

In addition to SHAP, this study also uses permutation feature importance to identify the most influential features on model performance. Fisher et al. (2019, pp. 1–2) explain that variable importance can be used to understand how much a variable contributes to model accuracy, including through permutation-based approaches. Vaarma and Li (2024, p. 6) also use permutation importance to identify features that contribute most to dropout prediction. With this approach, model results can be translated into information that is easier to understand and more useful for academic intervention.

Previous studies show that dropout prediction models continue to develop, but there remains a need to connect model results with academic actions that institutions can implement. Tete et al. (2022, pp. 1–2) found that the dropout prediction literature still rarely

reports or proposes managerial actions and educational policies after prediction models are developed. Therefore, this study not only compares the performance of Logistic Regression, Random Forest, and XGBoost but also adds McNemar's test and explainable machine learning so that prediction results can be interpreted more carefully and translated into a basis for academic intervention.

Based on the above discussion, this study aims to develop an explainable machine learning model to predict student dropout risk in vocational higher education. The dataset used is Predicting Student Dropout and Academic Success, which was developed from student data at higher education institutions in Portugal. Realinho et al. (2022, p. 2) explain that the dataset consists of 4,424 records and 35 attributes with no missing values and can be used for benchmarking the performance of various algorithms in dropout prediction and academic success problems. By using Logistic Regression, Random Forest, XGBoost, McNemar's test, permutation feature importance, and SHAP, this study is expected to contribute to the development of an early detection system that is not only accurate but also transparent and explainable.

RESEARCH METHOD

This study uses a quantitative approach with a computational experiment design based on machine learning. The focus of this study is to build, compare, and interpret student dropout risk prediction models in vocational higher education. The research stages include dataset collection, data preprocessing, target transformation, data splitting, model training, performance evaluation, statistical testing, and model interpretation using explainable machine learning. This design is in line with dropout prediction studies that use academic, demographic, socioeconomic, and early-semester performance data to build student risk classification models (Martins et al., 2023, p. 1; Realinho et al., 2022, pp. 1–2).

The study used the Predict Students' Dropout and Academic Success dataset obtained from the UCI Machine Learning Repository. The dataset contains student-level information derived from higher education institutions in Portugal and includes demographic, socioeconomic, macroeconomic, enrollment-related, and academic performance variables. In this study, the dataset was treated as a secondary open dataset; therefore, no additional data were collected directly from students (Realinho et al., 2022, p. 1).

The dataset consists of 4,424 student records and 35 attributes, with no missing values. The attributes in the dataset are grouped into demographic data, socioeconomic data, macroeconomic data, enrollment academic data, first-semester academic data, second-semester academic data, and classification target (Martins et al., 2023, pp. 2–3). The original target variable consisted of three outcomes: Dropout, Enrolled, and Graduate. To align the analysis with the purpose of early dropout risk identification, the target was reformulated into a binary classification task. Students in the Dropout category were coded as 1, while students in the

Graduate and Enrolled categories were grouped into the Non-dropout category and coded as 0. This transformation allowed the models to focus on distinguishing students at risk of dropout from students who were not categorized as dropout cases.

Data preprocessing was conducted by examining the feature structure, ensuring target availability, and converting the target into a binary format. Since the dataset had no missing values, imputation was not performed. After target transformation, the data were divided into training and testing sets using an 80:20 ratio with a stratified split. This technique was used to maintain the proportion of dropout and non-dropout classes in both the training and testing sets. Data splitting and evaluation using test data are commonly used in machine learning-based dropout prediction studies, particularly to maintain separation between the model training and testing processes (Vaarma and Li, 2024, p. 6).

Three models were used in this study: Logistic Regression, Random Forest, and XGBoost. Logistic Regression was used as a baseline because it is simple and commonly used in binary classification. Random Forest was used as a decision tree-based ensemble model, while XGBoost was used as a tree-based boosting model. The selection of these models was based on previous studies showing that algorithms such as Decision Tree, Support Vector Machine, Random Forest, XGBoost, CatBoost, and LightGBM are widely used in predicting student dropout and academic success (Villar and de Andrade, 2024, pp. 1–2). Martins et al. (2023, p. 3) also show that Decision Tree, Support Vector Machine, Naive Bayes, Gradient Boosting, Random Forest, and Artificial Neural Networks are widely used in predicting student academic performance and dropout.

The modeling implementation was carried out using Python. Logistic Regression was trained with feature standardization and balanced class weights. Random Forest was trained with 300 estimators and balanced class weights. XGBoost was trained with 300 estimators, a learning rate of 0.05, max depth of 4, subsample of 0.9, colsample bytree of 0.9, and scale positive weight to adjust for class imbalance. XGBoost was selected because it is a scalable tree boosting system designed to produce strong predictive performance across various machine learning tasks (Chen and Guestrin, 2016, p. 785). Villar and de Andrade (2024, pp. 4–5) also explain that boosting models, such as Gradient Boosting, XGBoost, CatBoost, and LightGBM, are used to improve classification ability and understand factors contributing to model output.

Model performance was examined from both predictive and statistical perspectives. Predictive performance was assessed using accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, and the confusion matrix. Because the dropout class is the main concern of this study, recall and F1-score were emphasized alongside accuracy. Vaarma and Li (2024, p. 6) emphasize that accuracy can be misleading when data are imbalanced because a model can appear to perform well by simply predicting the majority class; therefore, precision, recall, F1-

score, and the precision-recall curve are important for the dropout class.

In addition, McNemar's test was used to compare Logistic Regression and the best-performing descriptive model on the same test set. This test was included to determine whether the observed performance difference reflected a statistically meaningful improvement rather than only a numerical difference in evaluation metrics. The significance level used was 0.05. If the p-value is less than 0.05, the performance difference between the two models is considered statistically significant. The use of statistical testing in this study was intended to ensure that model comparison was not based solely on differences in descriptive metrics. This is in line with Goran et al. (2024, p. 122393), who emphasize that experimental results need to be validated through statistical methods to strengthen the reliability of conclusions.

Model interpretation was conducted using permutation feature importance and SHAP. Permutation feature importance was used to identify the features that most influenced model performance by randomly shuffling the values of a feature and observing the decrease in prediction score. Vaarma and Li (2024, p. 6) explain that permutation importance shows a decrease in model performance score when feature values are shuffled; the greater the decrease, the more important the feature is to the model. Nagy and Molontay (2024, p. 280) also explain that permutation importance works by shuffling feature values and observing their effect on model performance. In this study, permutation feature importance was calculated using the F1-score so that feature interpretation remained focused on the model's ability to identify the dropout class.

SHAP was used to explain the contribution of each feature to the model output, both in terms of feature importance and the direction of feature influence on dropout risk prediction. Nagy and Molontay (2024, pp. 277–281) explain that SHAP is based on the Shapley value concept and can be used to explain individual predictions and feature effects on model output. In the educational context, SHAP helps make complex models more transparent so that prediction results can be understood by stakeholders. This approach was used so that model results were not limited to risk predictions but could also be translated into more targeted academic intervention recommendations.

RESULTS AND DISCUSSION

Results

The original target distribution of the dataset consisted of three classes: Graduate, Dropout, and Enrolled. After transformation into a binary classification problem, the Graduate and Enrolled classes were combined as Non-dropout, while the Dropout class was retained as the risk class. The distribution of dropout risk classes is presented in Table 1.

Table 1. Distribution of Dropout Risk Classes

Class	Frequency	Percentage
Non-dropout	3,003	67,88
Dropout	1,421	32,12
Total	4,424	100,00

Table 1 shows that the Non-dropout class is larger than the Dropout class. The Non-dropout class consists of 3,003 records or

67.88%, while the Dropout class consists of 1,421 records or 32.12%. This proportion indicates that the dropout class is not dominant but remains large enough to be analyzed as a risk prediction problem.

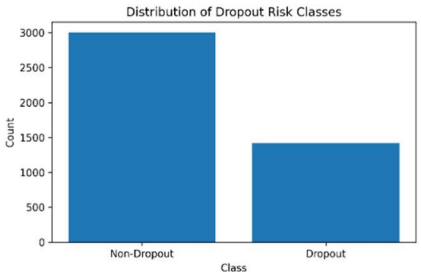

Figure 1. Distribution of Dropout Risk Classes

Three machine learning models were evaluated using the same test data. The model evaluation results are presented in Table 2.

Table 2. Performance Comparison of Dropout Risk Prediction Models

Metric	XGBoost	Logistic Regression	Random Forest
Accuracy	0,8859	0,8734	0,8847
Precision	0,8123	0,7867	0,8824
Recall	0,838	0,831	0,7394
F1-score	0,825	0,8082	0,8046
ROC-AUC	0,9355	0,9272	0,9316
PR-AUC	0,908	0,8948	0,9007

The comparison in Table 2 indicates that XGBoost produced the most favorable overall descriptive results among the three evaluated models, with an accuracy of 0.8859, an F1-score of 0.8250, an ROC-AUC of 0.9355, and a PR-AUC of 0.9080. Logistic Regression ranked second based on F1-score, while Random Forest had the highest precision but lower recall. This indicates that Random Forest was more selective in predicting dropout students but missed more students who actually dropped out compared to XGBoost and Logistic Regression.

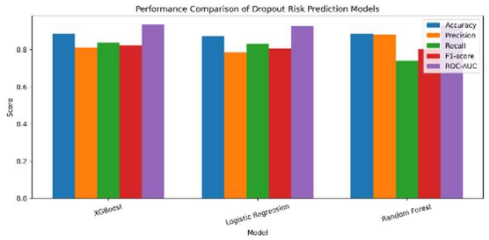

Figure 2. Performance Comparison of Dropout Risk Prediction Models

Based on the evaluation results, XGBoost was selected as the best descriptive model because it had the best balance between precision and recall for the dropout class. In the context of dropout prediction, recall is an important metric because failure to identify at-risk students may cause institutions to provide interventions too late. However, precision is also necessary so that institutions do not label too many students as at risk when they are actually not at risk. The performance of XGBoost for each class is presented in Table 3.

Table 3 shows that XGBoost classified the Non-dropout class with an F1-score of 0.9153 and the Dropout class with an F1-score of 0.8250. The recall value for the Dropout class was 0.8380, indicating that the model was able to identify most students who actually dropped out. This value is important because the main purpose of the model is to support early detection of at-risk students.

Table 3. Classification Report of the XGBoost Model

Class	**Prec.**	**Rec.**	**F1**	**N**
Non-dropout	0,9223	0,9085	0,9153	601
Dropout	0,8123	0,8380	0,8250	284
Macro avg	0,8673	0,8733	0,8701	885
Weighted avg	0,8870	0,8859	0,8863	885

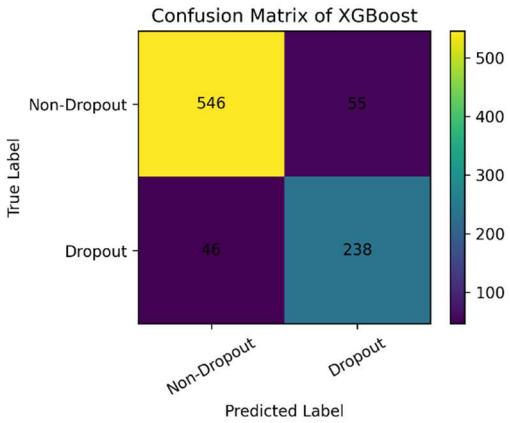

Figure 3. Confusion Matrix of the XGBoost Model

The confusion matrix provides a more detailed view of the prediction outcomes by separating correct classifications from misclassified cases 546 Non-dropout students and 238 Dropout students. Model errors consisted of 55 Non-dropout records predicted as Dropout and 46 Dropout records predicted as Non-dropout. In the context of academic intervention, false negatives need special attention because students who are actually at risk may not be detected by the system.

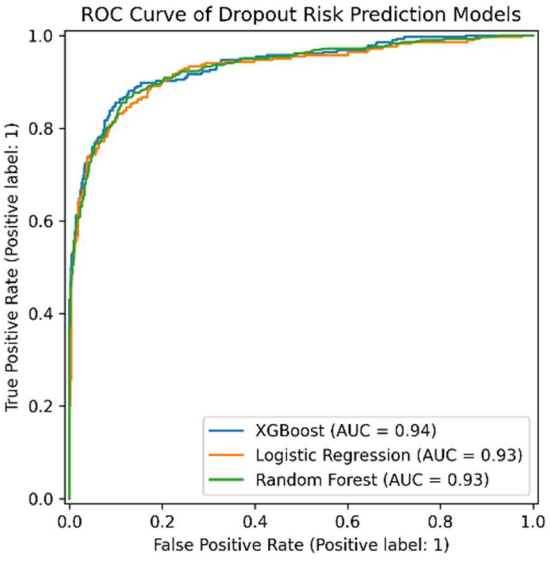

Figure 4. ROC Curve of Dropout Risk Prediction Models

The ROC curve shows that the three models had good discriminative ability. XGBoost obtained the highest ROC-AUC score of 0.9355, followed by Random Forest with 0.9316 and Logistic Regression with 0.9272. The differences in ROC-AUC values among the models were relatively small, so the advantage of XGBoost over the other models should be interpreted carefully. Therefore, this study added McNemar's test to examine whether the difference between XGBoost and the baseline model was statistically significant.

McNemar's test was conducted to compare Logistic Regression as the baseline model and XGBoost as the best descriptive model. The test results are presented in Table 4.

Table 4 shows that XGBoost corrected 32 predictions that were incorrect in Logistic Regression, while Logistic Regression corrected 21 predictions that were incorrect in XGBoost. Although XGBoost had higher descriptive performance, the exact p-value of 0.1690 was greater than 0.05.

Table 4. McNemar's Test Results between Logistic Regression and XGBoost

Item	Value
Both models correct	752
Logistic Regression correct, XGBoost incorrect	21
Logistic Regression incorrect, XGBoost correct	32
Both models incorrect	80
Discordant pairs	53
McNemar chi-square statistic	1,8868
McNemar chi-square *p-value*	0,1696
Exact McNemar *p-value*	0,1690
Interpretation	Not statistically significant

Therefore, the difference between Logistic Regression and XGBoost was not statistically significant. This finding indicates that XGBoost can be considered the best model descriptively, but it cannot yet be stated as significantly superior to the baseline model.

Permutation feature importance was used to identify the features that most influenced the predictive performance of XGBoost. The ten most important features are presented in Table 5.

Table 5 shows that second-semester academic features, particularly the number of approved curricular units, were the most dominant factors in predicting dropout risk. The feature Curricular units 2nd sem (approved) had the highest importance value of 0.2292.

The next most important features were Tuition fees up to date with 0.0672 and Curricular units 1st sem (approved) with 0.0264. These results indicate that early academic achievement and tuition fee payment status are important indicators in predicting dropout risk.

The SHAP beeswarm plot shows that the feature Curricular units 2nd sem (approved) had the strongest influence on the model output. High feature values tended to lead to negative SHAP values, which means that they reduced dropout risk. Conversely, low values of this feature tended to increase dropout risk.

The feature Tuition fees up to date also made a substantial contribution. Students with good payment status tended to have a lower dropout risk, while payment problems were associated with increased risk. Other features such as Curricular units 1st sem (approved), Age at enrollment, Course, and Debtor also contributed to the model prediction.

Table 5. Ten Most Important Features Based on Permutation Feature Importance

No	Feature	*Importance*
1	Curricular units 2nd sem (approved)	0,2292
2	Tuition fees up to date	0,0672
3	Curricular units 1st sem (approved)	0,0264
4	Curricular units 2nd sem (enrolled)	0,0221
5	Curricular units 2nd sem (grade)	0,0130
6	Curricular units 2nd sem (credited)	0,0116
7	Curricular units 1st sem (credited)	0,0093
8	Age at enrollment	0,0083
9	Debtor	0,0082
10	Gender	0,0073

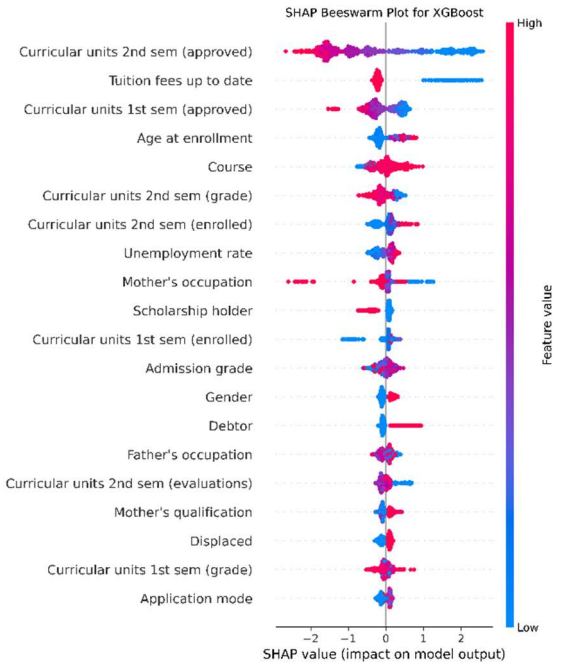

Figure 5. SHAP Beeswarm Plot of the XGBoost Model

Discussion

The results show that machine learning can be used to support early detection of dropout risk among vocational higher education students. XGBoost obtained the best descriptive

performance compared to Logistic Regression and Random Forest. This performance can be explained by the fact that XGBoost is a scalable tree boosting system designed to produce strong predictive performance across various machine learning tasks (Chen and Guestrin, 2016, p. 785). This finding is also in line with Villar and de Andrade (2024, pp. 1–2) who show that boosting-based models, such as XGBoost, CatBoost, and LightGBM, are relevant for predicting student dropout and academic success.

Although XGBoost had the best descriptive performance, McNemar's test showed that its difference from Logistic Regression was not statistically significant. This finding is important because it indicates that a more complex model does not always produce a sufficiently strong improvement over a simpler model. In the educational context, this result suggests that model selection should not be based solely on the highest accuracy score. Institutions also need to consider statistical significance, ease of implementation, interpretability, and resource readiness. Nagy and Molontay (2024, p. 280) show that complex models such as gradient-boosted trees may have higher performance than simpler models such as Logistic Regression, but interpretability remains an important aspect in dropout prediction. In addition, Goran et al. (2024, p. 122393) emphasize that experimental results need to be validated through statistical methods to strengthen the reliability of conclusions.

The model interpretation results show that early-semester academic performance was the most dominant factor in predicting dropout risk. The number of second-semester approved curricular units had the greatest influence, followed by the number of first-semester approved curricular units. This finding is consistent with Realinho et al. (2022, pp. 2–3), who explain that dropout prediction datasets contain academic features at the end of the first and second semesters and can be used to predict dropout and student academic success. Vaarma and Li (2024, p. 1) also show that transcript, demographic, and LMS activity data can be used to predict dropout in higher education. Thus, early academic achievement can be understood as an important indicator for detecting students' study continuity risk.

Financial-related variables also contributed to the prediction results. Tuition fee payment status appeared as one of the most influential features, suggesting that academic risk and financial stability may interact in shaping students' study continuity. In vocational higher education, students may face not only tuition obligations but also practical training expenses and other study-related costs. For this reason, payment status and debtor information should be viewed as early administrative signals that may help institutions provide timely support, not as a basis for labeling students negatively. Martins et al. (2023, p. 1) explain that academic difficulties, financial problems, personal issues, and family conditions can affect students' academic performance and contribute to dropout. Rahmani et al. (2024, p. 1) also show that financial constraints, motivation, support,

technology, and academic workload are factors that can affect dropout in higher education.

The use of explainable machine learning strengthens the practical value of the prediction model. Instead of producing only a risk classification, the model also provides information about which variables contribute most to the prediction. This is important in educational settings because academic staff need interpretable evidence before deciding what type of intervention should be offered. Permutation feature importance provides a global ranking of influential variables, while SHAP helps clarify how feature values contribute to the direction of model output. This approach is in line with Nagy and Molontay (2024, pp. 277–281), who explain that SHAP can be used to explain individual predictions and feature effects on model output. Lundberg and Lee (2017, p. 1) also explain that SHAP provides the contribution value of each feature to the model prediction, allowing complex model outputs to be understood.

Practically, the results of this study can be used as a basis for developing early warning systems in vocational higher education. Institutions can monitor early-semester academic indicators, payment status, and other risk indicators to determine which students need academic support. However, predictive models should still be positioned as decision-support tools, not replacements for academic judgment. Interpretation by academic advisors, study program managers, and student service units remains necessary so that interventions are not mechanical. Tete et al. (2022, pp. 1–2) show that many dropout prediction studies still lack a connection between model results and managerial actions or educational policies. Therefore, the explainable machine learning approach in this study can support more transparent, data-driven decision-making while still considering the human aspects of education.

Limitations

This study has several limitations. First, the dataset used in this study was collected from higher education institutions in Portugal. Therefore, the findings may not fully represent the characteristics of vocational higher education students in Indonesia, where academic systems, student support mechanisms, socioeconomic conditions, and institutional policies may differ. Second, model evaluation was conducted using a single stratified train-test split. Although this approach provides an initial estimate of model performance, it does not capture the stability of model performance across different data partitions.

Future studies should use cross-validation and, if possible, external validation using institutional data from Indonesian vocational higher education contexts. Third, this study focuses on predictive performance and model explainability using available dataset features.

Therefore, qualitative aspects, such as students' motivation, learning experiences, family background, and institutional support practices, were not examined directly. Future research can integrate predictive modeling with qualitative or institutional data to obtain a more comprehensive understanding of dropout risk.

CONCLUSION AND RECOMMENDATIONS

Conclusion

The findings suggest that machine learning can support the identification of students who may require early academic attention in vocational higher education. Among the three evaluated models, XGBoost achieved the most favorable descriptive results. This outcome is consistent with the characteristics of boosting-based tree models, which are capable of modelling complex interactions among tabular features. However, the statistical comparison shows that the advantage of XGBoost over Logistic Regression should be interpreted cautiously because McNemar's test did not indicate a significant difference between the two models.

The results show that XGBoost obtained the best descriptive performance, with an accuracy of 0.8859, an F1-score of 0.8250 for the Dropout class, an ROC-AUC of 0.9355, and a PR-AUC of 0.9080. However, McNemar's test showed that the difference between XGBoost and Logistic Regression was not statistically significant, with an exact p-value of 0.1690. Therefore, XGBoost can be considered the best model descriptively, but it has not been proven to be significantly superior to the baseline model.

The model interpretation results show that the most influential features were Curricular units 2nd sem (approved), Tuition fees up to date, and Curricular units 1st sem (approved). These findings indicate that early-semester academic performance and students' financial conditions are important indicators in predicting dropout risk. Therefore, explainable machine learning can be used to support early detection and the development of more targeted academic interventions in vocational higher education.

Recommendations

Future studies are recommended to use institutional data from vocational higher education institutions in Indonesia so that prediction results become more relevant to the local context. In addition, further studies can compare more models such as LightGBM, CatBoost, Support Vector Machine, and neural networks. Model testing should also be expanded using cross-validation or external validation to obtain a more stable estimate of model performance.

From an implementation perspective, the model results should be developed into an early warning system prototype that can be used by study programs or academic units. Such a system needs to consider ethical aspects, student data privacy, and the role of humans in decision-making. Machine learning predictions should be used as a tool to initiate academic support, not as the sole basis for evaluating students.

AUTHOR CONTRIBUTIONS

Author 1 was responsible for developing the research concept, designing the methodology, conducting the data analysis, preparing the initial manuscript, and revising the article. Author 2 supported data curation, literature mapping, and manuscript revision. Author 3 contributed to validation, literature review, and manuscript evaluation. Author 4 supported methodological

refinement, visualization, and technical checking. Author 5 provided supervision, validation, and final review of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest in the research and writing of this article.

FUNDING

This research was not supported by any specific grant from public, commercial, or non-profit funding institutions. The study used a publicly accessible secondary dataset and did not involve direct interaction with human participants; therefore, ethical clearance was not required.

REFERENCES

Andrade-Girón, D., Sandivar-Rosas, J., Marín-Rodriguez, W., Susanibar-Ramirez, E., Toro-Dextre, E., Ausejo-Sanchez, J., … Angeles-Morales, J. (2023). Predicting student dropout based on machine learning and deep learning: A systematic review. *EAI Endorsed Transactions on Scalable Information Systems*, *10*(5), 1–11. https://doi.org/10.4108/eetsis.3586

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

de Oliveira, C. F., Sobral, S. R., Ferreira, M. J., & Moreira, F. (2021). How does learning analytics contribute to prevent students' dropout in higher education: A systematic literature review. *Big Data and Cognitive Computing*, *5*(4), 64. https://doi.org/10.3390/bdcc5040064

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1–81. https://www.jmlr.org/papers/v20/18-760.html

Goran, R., Jovanovic, L., Bacanin, N., Stankovic, M. S., Simic, V., Antonijevic, M., & Zivkovic, M. (2024). Identifying and understanding student dropouts using metaheuristic optimized classifiers and explainable artificial intelligence techniques. *IEEE Access*, *12*, 122377–122400. https://doi.org/10.1109/ACCESS.2024.3446653

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774. https://arxiv.org/abs/1705.07874

Martins, M. V., Baptista, L., Machado, J., & Realinho, V. (2023). Multi-class phased prediction of academic performance and dropout in higher education. *Applied Sciences*, *13*(8), 4702. https://doi.org/10.3390/app13084702

Nagy, M., & Molontay, R. (2024). Interpretable dropout prediction: Towards XAI-based personalized intervention. *International Journal of Artificial Intelligence in Education*, *34*(2), 274–300. https://doi.org/10.1007/s40593-023-00331-8

Nti, I. K., & Ramanayake, S. (2026). Explainable machine learning for student dropout prediction and tailored interventions in online personalized education. *Discover Artificial Intelligence*, *6*(1), 16. https://doi.org/10.1007/s44163-026-01016-6

Rahmani, A. M., Groot, W., & Rahmani, H. (2024). Dropout in online higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, *21*(1), 10. https://doi.org/10.1186/s41239-024-

00450-9

Realinho, V., Machado, J., Baptista, L., & Martins, M. V. (2022). Predicting student dropout and academic success. *Data*, *7*(11), 146. https://doi.org/10.3390/data7110146

Tete, M. F., Sousa, M. de M., de Santana, T. S., & Fellipe, S. (2022). Predictive models for higher education dropout: A systematic literature review. *Education Policy Analysis Archives*, *30*(149), 1–23. https://doi.org/10.14507/EPAA.30.6845

Vaarma, M., & Li, H. (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. *Technology in Society*, *76*, 102474. https://doi.org/10.1016/j.techsoc.2024.102474

Villar, A., & de Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Discover Artificial Intelligence*, *4*(1), 79. https://doi.org/10.1007/s44163-023-00079-z

Wydra-Somaggio, G. (2021). Early termination of vocational training: Dropout or stopout? *Empirical Research in Vocational Education and Training*, *13*(1), 9. https://doi.org/10.1186/s40461-021-00109-z

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 2. https://doi.org/10.1186/s40561-022-00192-z