RESEARCH PAPER

Prediction of Academic Risk among University Students Using Ensemble Algorithms and Neural Networks

Iván Dennys Soto Rodríguez[1*], Frida Rosa Coaquira Nina[2] Fanny Campomanes Murrugarra[3]

[1*]Universidad Nacional Agraria La Molina Lima, Perú https://orcid.org /0000-0002-4928-8362 ivans@lamolina.edu.pe
[2]Universidad Nacional Agraria La Molina Lima, Perú fcoaquira@lamolina.edu.pe https://orcid.org/0000-0001-6531-595X
[3]Universidad Nacional Agraria La Molina fanny.campomanes@gmail.com https://orcid.org/0009-0000-4520-9703

Abstract
Academic risk among university students can result in low performance, slow academic progress, and dropout, which is why it is important to be identified at an early stage to provide institutional intervention. In this work, ensemble learning algorithms and neural networks were compared for the prediction of academic risk among students of Universidad Nacional Agraria La Molina. The academic records of 17,233 students entering from the first semester of 2010 to the first semester of 2023 were used for the study. Academic risk was defined as two groups: students with risk and students without risk. The predictor variables were enrolled credits, approved credits, failed credits, enrolled courses, approved courses, failed courses, weighted grade, career, faculty, and performance level. The original distribution was imbalanced, and oversampling was used prior to training the model. The models evaluated were CART, Bagging, Random Forest, XGBoost, and Neural Networks. The performance was evaluated using test data and five-fold cross-validation with accuracy, F-measure, PR-AUC, recall, ROC-AUC and specificity. The findings indicated that the most significant factors that predicted academic risk were weighted grade, failed credits, and failed courses. While Bagging had a slightly higher accuracy, the Neural Network had the highest overall performance with the highest recall, specificity, PR-AUC, and ROC-AUC in the test evaluation. The results showed that neural networks can be used to detect academic risk early and to assist in tutoring and academic counselling and in providing interventions for student welfare.

Keywords: Academic risk; neural networks; ensemble learning; student performance prediction; higher education; machine-learning

How to cite this article: Soto Rodríguez ID, Coaquira Nina FR, Campomanes Murrugarra F. Prediction of Academic Risk among University Students Using Ensemble Algorithms and Neural Networks. Int J Drug Deliv Technol. 2026;16(52s): 97-109. DOI: 10.25258/ijddt.16.52s.10

1. Introduction

The issue of university dropout is considered a significant problem for universities, as it impacts on students' progress, the performance of the institutions and the effective utilization of educational resources. Often, the causes of dropout are not single but are the result of a gradual process in which academic, personal, economic and social challenges build up over time. Of these, academic risk is especially important, because it involves the likelihood that a student will not reach the desired level of academic achievement, will be delayed in his/her progress through the curriculum, or will dropout of school altogether. Academic risk is thus related to indicators like low grades, failed courses, not earning sufficient credit, and not progressing academically. But it's important for universities to identify students who show these warning signs early in the process so they can take action to support the student in time before risk turns to failure, delay, or dropout.

Access in higher education is no longer just the goal, it is also expected that students are made permanent, achieve learning outcomes and complete their studies within a reasonable time. In this sense, universities need evidence-based mechanisms that can track the trajectory of students and detect academic vulnerability early on. When academic records are available at the institution, they provide an avenue to shift the paradigm from responding to academic issues to proactively intervening in them. In the past, a combination of administrative student data and machine-learning methods was found to be useful for early detection of students at risk (Berens et al., 2019). Predictive analytics has also been used to determine the likelihood that students in higher education are at risk for dropping out, further emphasizing the importance of data-driven approaches in higher education management (Gonzalez-Nucamendi et al., 2023).

Student permanence is a relevant issue in the university context in Peru, because of the disruptions faced before,

*Author for Correspondence: ivans@lamolina.edu.pe

during and after the COVID-19 pandemic. The figures for public universities revealed a variation in the number of entrants and graduates, including a significant drop during the pandemic. Before the pandemic, the number of participants in Universidad Nacional Agraria La Molina was consistent, with 736 participants in 2017, 720 in 2018 and 849 in 2019. However, the number of entrants fell to 641 in 2020, and the number of graduates dropped significantly to 178. In 2021, the number of entrants rose to 1013, but the number of graduates only grew to 537, which is still lower than before the pandemic. The patterns suggest the need to improve institutional policies and practices for retention, timely graduation, and academic support.

Machine-learning models can be useful tools to meet this need as they can detect complex patterns in student information and categorize students based on risk profiles. Research on university dropout prediction has shown that machine-learning algorithms can be applied to the analysis of academic records to enhance the classification of the risk of students dropping out of university (Cho et al., 2023). Comparative studies have also demonstrated that supervised machine-learning algorithms can be used to assist in predicting both dropout and academic success (Villar & de Andrade, 2024). Other studies have highlighted that student preferences and academic trajectories can also be important for predicting dropout, suggesting that the prediction of dropout is dependent on a few aspects of the learning experience (Segura et al., 2022). These predictive techniques have been applied to statistical models and machine-learning models for university dropout and scholarship-related outcomes (Romero & Liao, 2025) and to learning-management-system data for optimizing dropout prediction using these learning systems (Rebelo Marcolino et al., 2025). Machine-learning, deep-learning and explainable artificial intelligence are also emerging as increasingly relevant technologies in student-performance prediction, as confirmed by broader reviews (Boujmiraz et al., 2026).

Although academic records are available, universities need predictive models capable of identifying students at academic risk before such risk leads to failure, delay, or dropout. Therefore, this study compares ensemble algorithms and neural networks to determine which approach provides better predictive performance for academic-risk classification. Specifically, the study pursues the following objectives:

- To compare ensemble learning algorithms and neural networks in predicting academic risk among university students.
- To identify the variables with the greatest importance in the prediction of academic risk.
- To determine the most suitable predictive model for supporting early identification of students at academic risk.

2. Literature Review

2.1 Academic Risk Prediction in Higher Education

Academic risk prediction is an emerging field in educational data mining due to the need of higher education institutions to provide timely mechanisms for the identification of students at risk of academic failure, delayed progression, failure or dropout. Predictive modelling allows universities to identify patterns from the data they have on hand that can inform academic vulnerability and early intervention. The performance of students has been predicted using various computational techniques such as decision trees (Hamsa et al., 2016), fuzzy genetic algorithms (Hamsa et al., 2016), supervised machine-learning models, and other data-mining techniques. The methods demonstrate how educational information can be used to create usable indicators for academic monitoring.

The data that are typically used in studies on student-performance prediction vary based on what is available in each educational setting, including academic, demographic, behavioural, and institutional data (Albreiki et al., 2021). Academic variables are particularly significant since they directly relate to student progress, performance, and learning outcomes. Demographic and institutional factors can also help to explain student risk, however. Systematic reviews have also indicated that machine-learning models play a significant role in predicting academic performance by enabling researchers to compare algorithms, determine which predictors are relevant and develop more accurate classifications (Balaji et al., 2021). Educational data mining can thus serve as a structured approach to identify students at risk and evidence-based decision making in higher education (Zhang et al., 2021). In this context, early prediction is helpful in that it enables institutions to act before academic risks have become critical. The academic performance of students has been predicted using machine-learning algorithms and learners who might need support to achieve the final academic outcomes have been identified (Yağcı, 2022).

2.2 Ensemble Algorithms in Academic Prediction

Ensemble algorithms are effective to use in academic prediction as they combine several learners to enhance predictive performance, stability and generalization. An ensemble model combines multiple classifiers to produce a prediction that is more accurate and robust than a single model. Bagging is the combination of multiple models trained on bootstrap samples, which can be used for variance reduction. Random Forest is an extension of this idea that grows several decision trees and adds some randomness in the selection of predictors, enhancing the diversity of the models and providing variable-importance analysis. XGBoost works by boosting; that is, trees are added in sequence to rectify the errors of the ones before them. This allows it to be used to model more complex and non-linear patterns in student data. CART is one decision tree model that yields interpretable rules and can be used as a baseline model to compare with ensemble models, but it might not be as stable as ensemble models.

Recent research indicates that ensemble models are still useful for student academic prediction. In the context of academic-success prediction, ensemble learning has been utilized to fuse multiple modelling strategies and boost the classification accuracy (Zhao et al., 2025). Optimized ensemble deep-learning models have also been applied to predictive analysis of student achievement, demonstrating the potential of hybrid deep-learning and ensemble models in the field of education (Wang, 2024).

2.3 Neural Networks in Academic Prediction

Neural networks are predictive models which are nonlinear and can learn complex relationships between variables. This is important for academic-risk prediction because student outcomes might be influenced by several academic factors including performance level, failed courses, grades, and credits. A neural network is a set of nodes in layers that are connected to each other, receive input variables, and during training, adjust the parameters to reduce the difference between the predicted and the actual values. They can capture the nonlinear relationships, making them suitable for a classification problem in which academic risk is influenced by multiple factors.

Neural and deep-learning models have been applied in recent prediction studies in the field of education to enhance the classification of academic performance. Neural architectures are important in educational prediction, and in the field of student academic achievement prediction, stacked ensemble learning with deep neural networks and fuzzy-based feature selection has proven to be effective in educational prediction tasks (Gu, 2025). Likewise, deep-learning models based on Bi-LSTM, which are interpretable and statistically validated, have been employed to predict academic performance, demonstrating the ability of neural networks to predict and explain student performance in academic settings (Kalita et al., 2025). The neural network should generally be carefully tuned, and it requires more computation to run than simpler models, so the performance should be compared with other algorithms with a few metrics.

2.4 Research Position of the Present Study

The reviewed studies show that both ensemble algorithms and neural networks are useful for predicting academic performance and identifying students at risk. Ensemble models offer stability, robustness, and variable-importance analysis, while neural networks provide strong capacity to model complex nonlinear relationships. However, model performance depends on the dataset, predictor variables, class distribution, preprocessing method, and validation strategy. Therefore, this study evaluates ensemble algorithms and neural networks using institutional academic data from UNALM to determine the most suitable model for academic-risk prediction.

3. Materials and Methods

3.1 Study Design

The study followed an applied, quantitative, predictive, longitudinal and quasi-experimental design. It was applied because it was designed to create a functional model to identify students at academic risk in university. It was quantitative because it was based on academic records and measurable indicators of student performance that were structured. The study was predictive since the goal was to classify students based on their academic-risk condition, with the help of machine-learning algorithms. The longitudinal nature of the study was determined by the use of institutional records of several academic periods (first semester of 2010 to first semester of 2023). The design was quasi-experimental as the analysis was based on the existing institutional data and no random selection of students to experimental groups.

3.2 Data Source and Study Population

The study data consisted of academic records from Universidad Nacional Agraria La Molina. These records were collected from the Office of Academic Records and comprised students from all specialities who enrolled in the university from the first semester of 2010 to the first semester of 2023. The study population was 17,233 students and the analysis was carried out with the entire available population.

Academic risk was the response variable, with two categories: Students with academic risk and students without academic risk. The original distribution had 80% of the population (13,786) without academic risk and 20% of the population (3,447) with academic risk. This distribution showed a clear class imbalance, and this was considered when building the model.

3.3 Study Variables

Academic risk was the dependent variable and was defined as the likelihood that a student will demonstrate poor academic performance that may result in academic failure or dropout. Academic load, academic achievement, programme affiliation and performance level were the predictor variables. These variables encompassed enrolled credits, approved credits, failed credits, enrolled courses, approved courses, failed courses, weighted grade, career, faculty, and academic performance. Performance was rated as high, medium and low.

These variables were chosen as they reflect the student's academic pathway and give information about academic progress and academic challenge. Approved and failed credits approved and failed courses, and weighted grade reflect the performance of the student, whereas career and faculty reflect the institutional context.

3.4 Data Preprocessing and Balancing

An exploratory analysis was first performed to explore the distribution of the study variables and the academic-risk categories. The analysis revealed that the data set was imbalanced with more students with no academic risk than students with academic risk. This imbalance might result in a model favouring the majority class and not being as effective at identifying the minority class, as the goal was to identify students at risk.

To correct this, the oversampling technique was used to equalize the risk and non-risk groups. This method boosted the number of students with academic risk in the training of the models and enabled the algorithms to better learn from the minority class. In the cross-validation, the folds were oversampled only in the training folds, and the validation fold was kept independent. This process ensured that information was not leaked, and the validity of model evaluation was maintained.

3.5 Predictive Models and Validation

Five supervised machine-learning models were tested: CART, Bagging, Random Forest, XGBoost, and Neural Network. A baseline tree-based classifier was CART. The ensemble learning techniques included in the study were bagging, random forest, and XGBoost, while the nonlinear relationships between the academic variables were modelled using the neural network.

The data set was split into 80% training and 20% testing. Furthermore, five-fold cross-validation was used to test the stability and generalization of the models. Four folds were used for training and one-fold for validation in each iteration of cross-validation. This was repeated until each fold had been used one time as the validation fold.

The hyperparameters were tuned for each model. CART employed cost complexity, tree depth and minimum node size. The cost complexity, tree depth and minimum node size were used as the cost functions in bagging. Random Forest used the number of predictors selected at each split, number of trees, minimum node size, and impurity-based variable importance. XGBoost employed number of predictors, number of trees, minimum node size and tree depth. The Neural Network comprised of a penalty, 200 epochs, ReLU activation, Adam optimizer, and a batch size of 32.

3.6 Model Evaluation

The accuracy, F-measure, precision-recall area under the curve, recall, receiver operating characteristic area under the curve, and specificity were used to evaluate the models. The accuracy was calculated as the overall percentage of correctly classified examples, and the F-measure gave an overview of the balance between the precision and the recall. The importance of recall was due to the goal of correctly identifying students with academic risk. Specificity was also taken into account as the model had to be able to accurately classify students who were not at academic risk. The quality of classification beyond a single threshold was evaluated using ROC-AUC and PR-AUC.

The final interpretation was done using several performance metrics, as the original data set was imbalanced and the accuracy was not the only metric considered. This method enabled more reliable comparison of the algorithms and helped select the model that had the highest overall ability to predict academic risk.

4. Results

4.1 Distribution of Academic Risk

The study encompassed 17,233 students in two academic-risk groups. Based on the data in Table 1, the total number of students identified in this study as without academic risk was 13,786 (80%) and the total number of students identified in this study as with academic risk was 3,447 (20%). This distribution shows that there is a definite imbalance between the two groups, with students at risk being the minority class. The same pattern is visually displayed in Figure 1, which shows that a significantly larger number of students are not at risk, compared to the number of students at risk.

Table 1. Distribution of students according to academic-risk condition

Academic-risk condition	Frequency	Percentage

Without risk	13,786	80%
With risk	3,447	20%
Total	17,233	100%

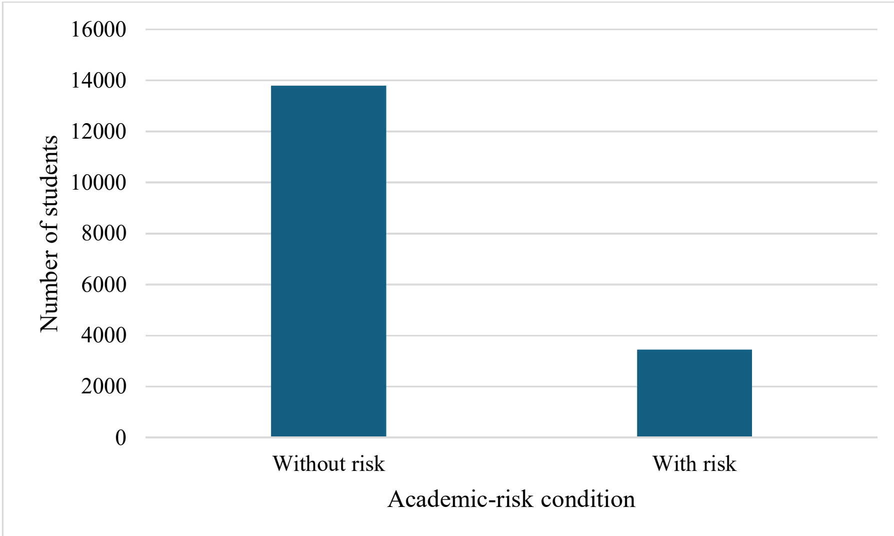

Figure 1. Distribution of students according to academic-risk condition.

Because the risk group represented only one-fifth (20%) of the study population, oversampling was applied prior to model training. As shown in Figure 2, the original distribution was imbalanced prior to oversampling; after oversampling, the numbers of students in the risk and non-risk classes were balanced. This step was necessary to reduce the model's tendency to favour the majority class and to improve its ability to identify students at academic risk.

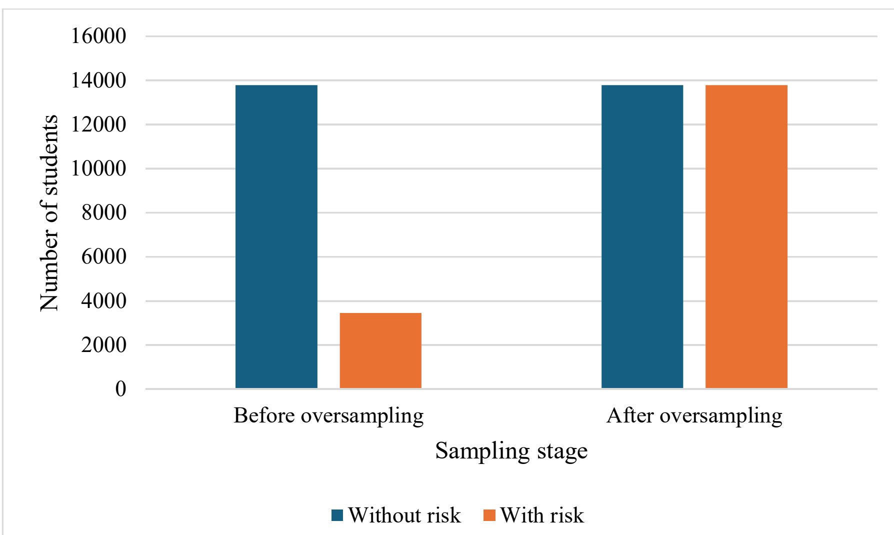

Figure 2. Distribution of academic-risk classes before and after oversampling.

4.2 Descriptive Analysis of Predictor Variables

The quantitative variables were described and differences in their distribution were found based on academic-risk condition. There was little separation between students with and without academic risk on enrolled credits or enrolled courses as seen in Figure 3. This means that the academic load students registered was not the most obvious distinction between the two groups.

However, there were more apparent differences between students with and without academic risk for approved credits, failed credits, approved courses, failed courses, and weighted grade. Students with academic risk scored higher in the categories of failed credits, failed courses, and students without academic risk scored higher in the categories of approved credits, approved courses, and weighted grade. The boxplots in Figure 3 also indicate that weighted grade was lower for students with academic risk, further demonstrating its value as an indicator of students' academic performance.

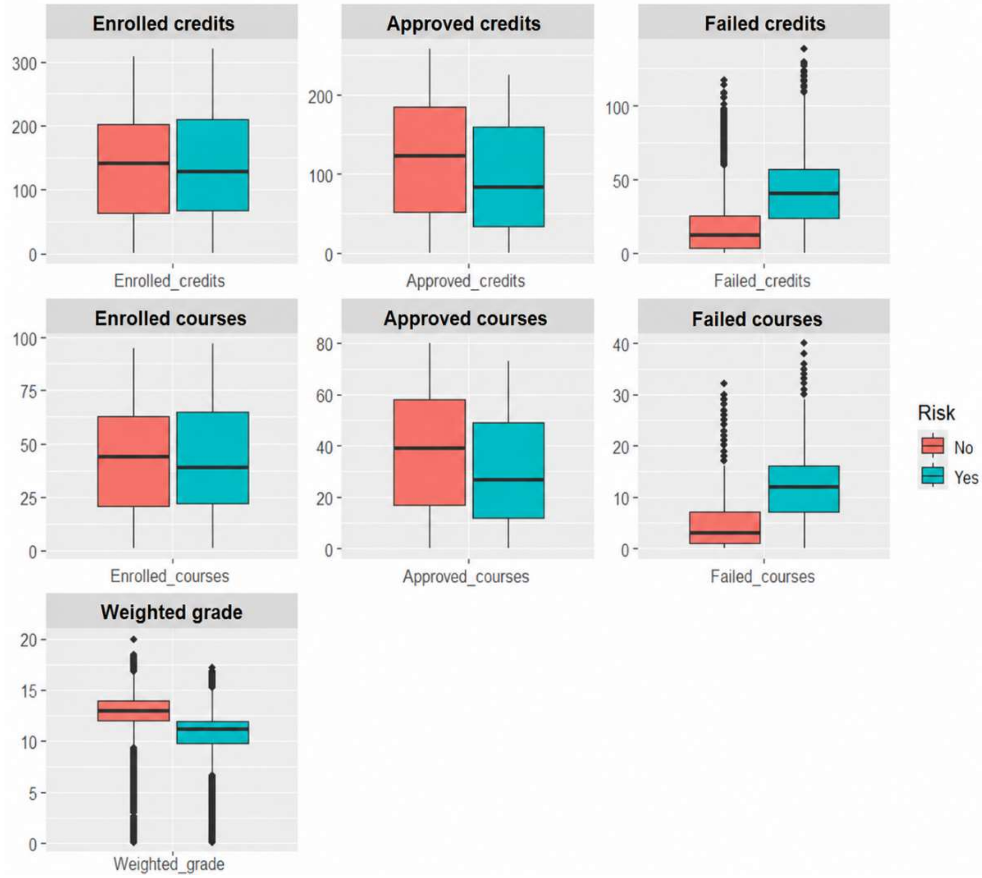

Figure 3. Distribution of quantitative predictor variables according to academic-risk condition.

The qualitative analysis revealed that the distribution of academic-risk was different at performance level, career and faculty level. The results revealed that the direct academic-performance indicators gave better differentiation between the risk categories than the programme level variables. The next model comparison thus involved testing the selected algorithms on the ability to classify academic risk using these predictors.

4.3 Variable Importance

Random Forest algorithm was used to determine the variable importance. The results showed that weighted grade was the most important predictor of academic risk. This means that the student's credit-weighted academic average was the best predictor of identifying students at risk and at no risk. The next most important predictors, summarized in Table 2, were failed credits and failed courses, which indicate a history of academic difficulty. Approved credits and approved courses were also relevant as they represented successful academic progress. Low performance was also found to be an important predictor.

Table 2. Main predictors of academic risk based on Random Forest importance

Rank	Predictor	Interpretation
1	Weighted grade	Strongest indicator of academic-risk condition
2	Failed credits	Reflects accumulated credit-level academic difficulty
3	Failed courses	Indicates course-level academic failure
4	Approved credits	Represents successful academic progression
5	Approved courses	Indicates successful course completion
6	Low performance	Reflects weak academic standing

The findings in Table 2 indicate that direct academic-performance variables were more informative than career and faculty. This suggests that the most useful predictors of academic risk were variables that directly measured student achievement, course completion, and academic failure.

4.4 Cross-Validation Performance

The performance of the five-fold cross-validation revealed that Bagging and Neural Network models were the most competitive models. The highest cross-validation accuracy was achieved by bagging and the Neural Network performed better in recall, specificity and ROC-AUC. This distinction is critical as the purpose of the study was to not only maximize the total number of correct classifications, but also to correctly identify students with academic risk.

The Neural Network achieved a cross-validation recall of 0.9117 as compared to Bagging, CART and XGBoost as shown in Table 5. It also achieved the highest specificity in cross-validation (0.7819), which is a better indicator of identification of students not at academic risk. CART performed poorly on most measures, and XGBoost performed well on some measures but not as well on specificity.

4.5 Test Data Performance

The test data results indicated that the performance of the models evaluated were similar in terms of accuracy and F-measure while there were significant differences in recall, PR-AUC, ROC-AUC and specificity. With the results shown in Table 3, Bagging achieved the highest accuracy (0.8141), followed by the Neural Network (0.8125) and XGBoost (0.8115). CART had the lowest accuracy (0.7831). The detailed test data and cross-validation performance metrics that were available for comparison were Bagging, CART, Neural Network and XGBoost. Random Forest was included in the ROC-AUC comparison of the classifiers because the result of the ROC curve was reported separately for this model in the model-discrimination analysis.

Table 3. Test data performance of the evaluated models

Algorithm	Accuracy	F-measure	PR-AUC	Recall	ROC-AUC	Specificity
Bagging	0.8141	0.8788	0.6048	0.8475	0.8565	0.6846
CART	0.7831	0.8519	0.5371	0.7848	0.8346	0.7765
Neural Network	0.8125	0.8790	0.6228	0.9163	0.8702	0.8105
XGBoost	0.8115	0.8806	0.5610	0.8745	0.8305	0.5672

The pattern shown in Table 3 is compared in Figure 4. As can be seen in the figure, the accuracy values of Bagging, Neural Network and XGBoost were very close. The Neural Network, however, had superior performance in recall, PR-AUC, ROC-AUC and specificity. The highest recall was obtained by the Neural Network with a value of 0.9163, which means that it has the best capability to recognize a student at academic risk. It also achieved the best specificity of 0.8105, which was better in discriminating students with no academic risk. Although XGBoost had the highest F-measure (0.8806), its specificity was the lowest (0.5672), indicating a greater tendency to misclassify students without academic risk.

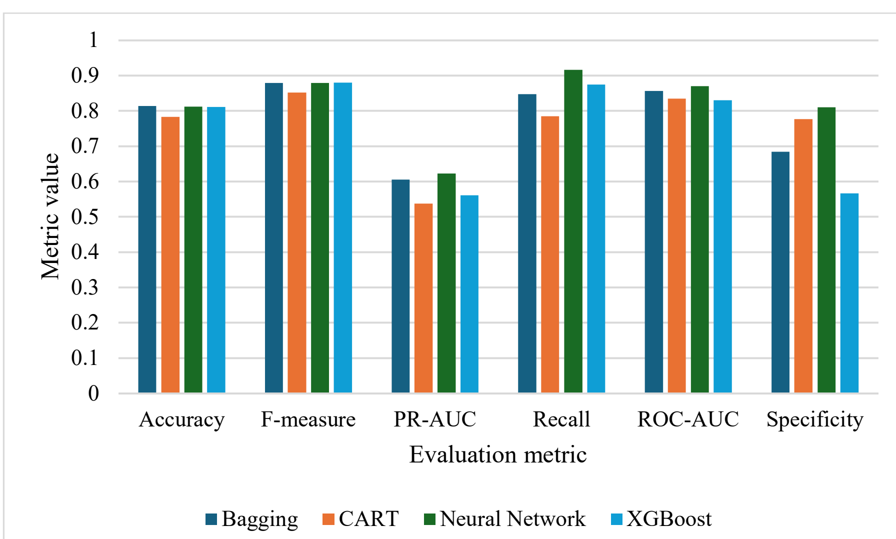

Figure 4. Test data performance metrics of the evaluated algorithms.

4.6 ROC-AUC Comparison

The ROC-AUC comparison also showed that the Neural Network had a higher discriminatory ability. The highest ROC-AUC value was achieved by the Neural Network with 0.870, followed by Bagging with 0.856 and Random Forest with 0.852 as shown in Table 4. The ROC-AUC values for CART and XGBoost were lower, 0.835 and 0.831, respectively.

Table 4. ROC-AUC comparison of the evaluated algorithms

Algorithm	ROC-AUC
Neural Network	0.870
Bagging	0.856
Random Forest	0.852
CART	0.835
XGBoost	0.831

The same pattern can be observed in Figure 5, which demonstrates the best model in terms of ROC-AUC is the Neural Network. This means the overall ability of the Neural Network to classify students with and without academic risk at different classification thresholds was the highest.

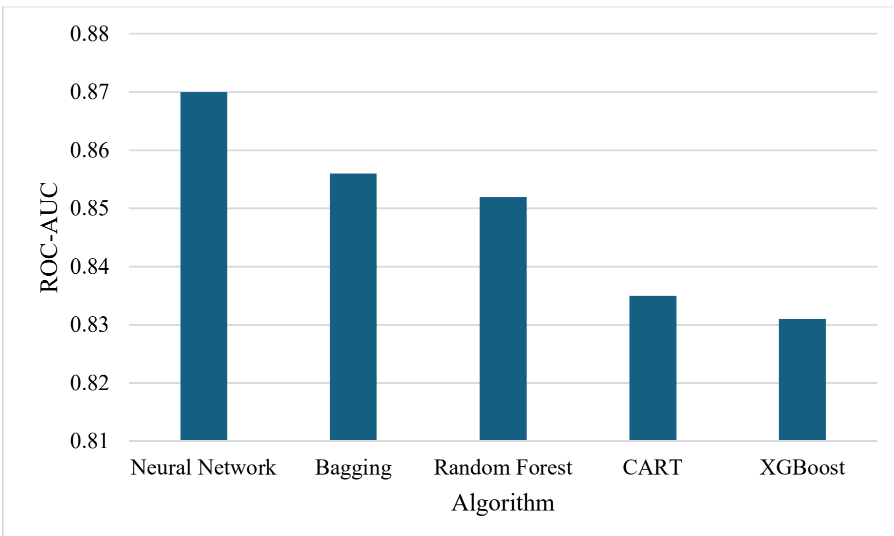

Figure 5. ROC-AUC comparison of the evaluated algorithms.

4.7 Test and Cross-Validation Comparison

Results of the comparison of test and cross-validation are shown in Table 5. The Neural Network performed well in both evaluation methods. It achieved the highest recall, PR-AUC, ROC-AUC and specificity in the test evaluation. For academic-risk classification, it also achieved the highest recall and specificity of the reported models in the cross-validation, which indicates its stability.

Table 5. Test and cross-validation performance comparison

Algorithm	Metric	Test value	Cross-validation value
Bagging	Accuracy	0.8141	0.8321
CART	Accuracy	0.7831	0.7942
Neural Network	Accuracy	0.8125	0.8230
XGBoost	Accuracy	0.8115	0.8170
Bagging	F-measure	0.8788	0.8948
CART	F-measure	0.8519	0.8623
Neural Network	F-measure	0.8790	0.8937

XGBoost	F-measure	0.8806	0.8847
Bagging	PR-AUC	0.6048	0.9488
CART	PR-AUC	0.5371	0.9166
Neural Network	PR-AUC	0.6228	0.9535
XGBoost	PR-AUC	0.5610	0.9379
Bagging	Recall	0.8475	0.8976
CART	Recall	0.7848	0.8113
Neural Network	Recall	0.9163	0.9117
XGBoost	Recall	0.8745	0.8831
Bagging	ROC-AUC	0.8565	0.8460
CART	ROC-AUC	0.8346	0.8246
Neural Network	ROC-AUC	0.8702	0.8567
XGBoost	ROC-AUC	0.8305	0.8237
Bagging	Specificity	0.6846	0.5781
CART	Specificity	0.7765	0.7278
Neural Network	Specificity	0.8105	0.7819
XGBoost	Specificity	0.5672	0.5607

As shown in Table 5, Bagging was competitive in accuracy and F-measure, but its specificity was lower than that of the Neural Network. CART showed weaker performance overall, particularly in accuracy, F-measure, PR-AUC, and recall. XGBoost obtained the highest test F-measure but had the lowest specificity, which limits its suitability when correct identification of students without academic risk is also required.

4.8 Final Model Selection

Based on the detailed performance metrics in Tables 3 and 5, the ROC-AUC comparison in Table 4 and Figure 5, the Neural Network model was chosen as the most appropriate model for academic-risk prediction. The highest accuracy was obtained by Bagging, but the Neural Network was the best overall classifier as it had the highest test data recall, specificity, PR-AUC and ROC-AUC. The high recall suggests it has a good ability to identify students at risk academically, whereas the high specificity suggests it has a good ability to identify students not at risk academically. Bagging and XGBoost were competitive in selected metrics, however, their specificity was less than that of other approaches, which made them less widely applicable. The worst performance was for CART and hence it was the least desirable model for the classification problem.

5. Discussion

The most important result of the study was that the Neural Network model was the most accurate on average in predicting academic risk when compared to the other models analyzed. Bagging gave the best accuracy, but accuracy was not enough to make it to the final model as the study was conducted on an imbalanced academic-risk structure. In this context, a model can be very accurate in correctly identifying the majority group and miss the identification of students who are really at risk. Thus, recall, specificity, ROC-AUC and PR-AUC were the areas that needed to be paid more attention to when interpreting the performance. Recall was especially relevant as the core purpose of academic-risk prediction is to identify students that need timely academic support. A low recall model would overlook more students at risk, thus decreasing the opportunity for early intervention. The Neural Network had the best recall of all the models, meaning that it was able to identify the most students at academic risk. This finding is important because it is crucial to identify students early on to prevent academic delay, poor academic progression, and potential dropout. In higher education, other similar research has focused on the need for predictive tools to detect risk in time for institutions to take action to prevent students from dropping out or failing to perform well (Martins et al., 2023).

Specificity was also crucial as it was important that the academic-risk prediction system not only found vulnerable students, but also not generate too many false alarms. Many students without risk may be misidentified as at risk, which might result in misallocation of institutional resources and/or inefficiency in support services. The most specific model was the Neural Network, which had the best of discriminating students that are not at risk of academic failure. This is the reason why the Neural Network is more appropriate for the institutional decision-making process than a model that is chosen based on its accuracy. The need for multiple evaluation metrics is similar to that of imbalanced student dropout prediction where using accuracy will lead to incorrect conclusions when one class dominates the data (Mduma, 2023).

The results from the variable importance analysis indicated that weighted grade was the most important predictor of academic risk. This is a logical result as it mirrors the actual achievement of students in the courses they have registered and is a composite of their weighted grade. Lower weighted grade could be a red flag for ongoing problems with academic standards and may be an indicator for early intervention. Academic-performance indicators' importance in student-risk prediction is also reinforced by studies indicating that academic-performance variables are highly informative in predicting academic outcomes and dropout in higher education (Won et al., 2023).

Other significant predictors were failed credits and failed courses. These variables are cumulative indicators of academic difficulty and suggest that students are not successfully completing a portion of their academic work. The significance of their importance indicates that a low grade is not the only indicator of academic risk and that repeated patterns of failure are also associated with academic risk. These are good indicators because they are typically found in institutional academic records and can be tracked over each academic term. The other approved credits and approved courses also helped to predict as they indicate academic progression and successful completion of courses. Both factors suggest that academic-risk classification is highly dependent on the academic failure-advancement linkage.

Direct academic-performance variables had more influence than were career and faculty. However, this does not imply that programme-level context is unimportant, but rather, that in this study immediate academic indicators were more informative. From a practical perspective, this means that student-level patterns of performance, such as weighted grade and course/credit outcomes should be emphasized in institutional monitoring systems. This interpretation is in line with the existing evidence that dropout and academic risk are multi-dimensional, and that one of the most salient dimensions that can be observed is academic performance (Barragán Moreno & González Támara, 2024).

The Bagging model was competitive and stable, especially in terms of accuracy and F-measure, as compared to other models. This suggests that ensemble methods using multiple decision trees are good classifiers in academic-risk prediction. The advantage of bagging might have been its ability to decrease the variance, by aggregating together several tree-based learners, making it less sensitive to the variations of a single decision tree. However, the lower specificity of this model (compared to the Neural Network) made it less suitable for the final model, when both risk detection and correct identification of non-risk students were considered.

XGBoost also had competitive values in some of the metrics like F-measure and recall with lower specificity value. This indicates that the model identified more students who were at risk for academic failure but was more likely to misclassify students who were not at risk for academic failure. This is an important constraint for an institution-based EWS system, as over-alarms could affect the accuracy of the allocation of support. The CART model was not as effective as the other models, suggesting that a single decision-tree structure was not enough to explain the complexity of academic-risk classification. This discovery is in line with the benefits of the more flexible modelling methods like ensemble models and neural networks.

The Neural Network needed the longest computational time but had the best overall predictive performance. The benefit of this is that it may be explained through the fact that it could model nonlinear relationships between academic variables. There are likely multiple factors that interact to influence academic risk such as weighted grade, failed credits, failed courses, and approved courses. More complex models are not as good as neural networks for capturing such interactions. Recent research using artificial intelligence (AI) techniques also suggests that sophisticated predictive models can help forecast student success in school and offer valuable insights for early identification of students at risk (Paul et al., 2025).

The results are relevant to the institutions. A predictive model that has a high level of recall and specificity can be used to help identify students at academic risk early, and then to enable universities to take action before academic issues result in failure, delayed progression or dropout. The chosen model of a Neural Network can be used as an academic monitoring decision support tool to identify students that might need tutoring, academic counselling or student welfare support. The findings also suggest that academic-performance indicators should be the main emphasis of institutional interventions. Weighted grade, failed credits and failed courses are good variables to monitor as risk indicators. This can assist tutoring services in identifying those students who are at risk of failing in the classroom and focus their attention on them. Student welfare offices might also also use such information to coordinate support when there is a connection between academic risk and other personal, social or economic issues. Similarly, machine-learning approaches to developing retention policies have highlighted the importance of predictive models to help inform institutional proactive retention policies instead of waiting for students to dropout (Hoca & Dimililer, 2025).

The study also contributes to the academic management based on data. While the universities already have a lot of information about their students' academic performance, this information could be underutilized unless it is converted into actionable risk indicators. Predictive modelling can be used to transform the existing, routine academic information into structured information to support timely decision making. Previous research with digital and behavioural learning data has also demonstrated that the analysis of such data can be used for student-performance prediction and learning assessment, albeit the scope of prediction depends on the type of data that is available (Borna et al., 2024).

The study has some limitations which have to be taken into consideration when interpreting the findings. First, the analysis was performed because of the data of one university, so the results are the academic structure, student population and the context of this university. Secondly, prediction was done using the variables that were available in the academic record. These were very informative variables, but other factors like socioeconomic conditions, attendance, psychological support, motivation or behaviour in the learning-platform were not included in the model. Third, only Bagging, Random Forest, XGBoost, CART and neural network models were evaluated. Other machine-learning or deep-learning methods can yield other results. Lastly, the Neural Network model needed the most computational time when compared to the other models evaluated. Thus, it had the best overall predictive performance, but it should be considered whether it is going to be used in an institutional system and what the computational resources and operational needs are.

6. Conclusion

The aim of this study was to compare the prediction of academic risk among university students using ensemble learning algorithms and Neural Networks. The models evaluated were Bagging, Random Forest, XGBoost, CART and Neural Networks. Oversampling was used prior to model training to balance the number of academic risk and non-academic risk students in the study population, and multiple metrics were used to interpret model performance, such as accuracy, F-measure, PR-AUC, recall, ROC-AUC, and specificity. The findings show that academic-risk prediction was mainly influenced by direct academic-performance indicators. Weighted grade was the most significant predictor, followed by failed credits and failed courses. The other approved credits and approved courses also had a significant influence on prediction, and career and faculty had a lesser influence. This suggests that student performance at the programme level is not as strongly correlated with academic risk as are the characteristics of student performance over time. The Neural Network had the strongest overall predictive performance of the models evaluated. Although Bagging obtained the highest accuracy, the Neural Network achieved the highest recall, specificity, PR-AUC, and ROC-AUC. As such, it is the most appropriate model for academic-risk detection since it was more effective in identifying students at risk but also differentiated students without risk. Bagging was also competitive, whereas CART showed the lowest performance. The above results suggest that the Neural Network model can be used as a decision-support model for early academic-risk identification. Academic monitoring systems should give due consideration to weighted grade, failed credits and failed courses. The model can assist institutions to identify the vulnerable students before academic risk results in failure, delay, or dropout, and thus facilitate tutoring, academic counselling and student welfare interventions. But implementation should take into consideration the greater computational time needed for neural networks and further test the model using future institutional records.

References

1. Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, *11*(9), 552.
2. Balaji, P., Alelyani, S., Qahmash, A., & Mohana, M. (2021). Contributions of machine learning models towards student academic performance prediction: A systematic review. *Applied Sciences*, *11*(21), 10007.
3. Barragán Moreno, S. P., & González Támara, L. (2024, November). Complexities of student dropout in higher education: A multidimensional analysis. In *Frontiers in Education* (Vol. 9, p. 1461650). Frontiers Media SA.
4. Berens, J., Schneider, K., Gortz, S., Oster, S., & Burghoff, J. (2019). Early Detection of Students at Risk--Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods. *Journal of Educational Data Mining*, *11*(3), 1-41.
5. Borna, M. R., Saadat, H., Hojjati, A. T., & Akbari, E. (2024, December). Analyzing click data with AI: implications for student performance prediction and learning assessment. In *Frontiers in education* (Vol. 9, p. 1421479). Frontiers Media SA.
6. Boujmiraz, S., Darhmaoui, H., & El Maliani, A. D. (2026). Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches. *Computers and Education: Artificial Intelligence*, 100548.

7. Cho, C. H., Yu, Y. W., & Kim, H. G. (2023). A study on dropout prediction for university students using machine learning. *Applied sciences*, *13*(21), 12004.
8. Gonzalez-Nucamendi, A., Noguez, J., Neri, L., Robledo-Rella, V., & García-Castelán, R. M. G. (2023, October). Predictive analytics study to determine undergraduate students at risk of dropout. In *Frontiers in Education* (Vol. 8, p. 1244686). Frontiers Media SA.
9. Gu, J. (2025). Predicting student academic achievement using stacked ensemble learning with deep neural networks and fuzzy-based feature selection. *Scientific Reports*, *15*(1), 37195.
10. Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. *Procedia technology*, *25*, 326-332.
11. Hoca, S., & Dimililer, N. (2025). A machine learning framework for student retention policy development: A case study. *Applied Sciences*, *15*(6), 2989.
12. Kalita, E., Alfarwan, A. M., El Aouifi, H., Kukkar, A., Hussain, S., Ali, T., & Gaftandzhieva, S. (2025, June). Predicting student academic performance using Bi-LSTM: a deep learning framework with SHAP-based interpretability and statistical validation. In Frontiers in Education (Vol. 10, p. 1581247). Frontiers.
13. 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.
14. Mduma, N. (2023). Data balancing techniques for predicting student dropout using machine learning. *Data*, *8*(3), 49.
15. Paul, R., Sarker, S., El Aouifi, H., Hussain, S., Baruah, A. K., & Gaftandzhieva, S. (2025, December). Analyzing dropout of students and an explainable prediction of academic performance utilizing artificial intelligence techniques. In *Frontiers in Education* (Vol. 10, p. 1698505). Frontiers Media SA.
16. Rebelo Marcolino, M., Reis Porto, T., Thompsen Primo, T., Targino, R., Ramos, V., Marques Queiroga, E., ... & Cechinel, C. (2025). Student dropout prediction through machine learning optimization: insights from moodle log data. *Scientific Reports*, *15*(1), 9840.
17. Romero, S., & Liao, X. (2025). Statistical and machine learning models for predicting university dropout and scholarship impact. *PLoS One*, *20*(6), e0325047.
18. Segura, M., Mello, J., & Hernández, A. (2022). Machine learning prediction of university student dropout: Does preference play a key role?. *Mathematics*, *10*(18), 3359.
19. 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), 2.
20. Wang, K. (2024). Optimized ensemble deep learning for predictive analysis of student achievement. *Plos one*, *19*(8), e0309141.
21. Won, H. S., Kim, M. J., Kim, D., Kim, H. S., & Kim, K. M. (2023). University student dropout prediction using pretrained language models. *Applied Sciences*, *13*(12), 7073.
22. Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.
23. Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: method review and comparison analysis. *Frontiers in psychology*, *12*, 698490.
24. Zhao, S., Zhou, D., Wang, H., Chen, D., & Yu, L. (2025). Enhancing student academic success prediction through ensemble learning and image-based behavioral data transformation. *Applied Sciences*, *15*(3), 1231.