Applied Computer Science, vol. 20, no. 3, pp. 67–84
doi: 10.35784/acs-2024-29

Submitted: 2024-04-05 | Revised: 2024-05-31 | Accepted: 2024-06-14

Keywords: classification, regression, student performance, Machine Learning

Bilal OWAIDAT [0000-0002-4176-2197]*

EXPLORING THE ACCURACY AND RELIABILITY OF MACHINE LEARNING APPROACHES FOR PREDICTING STUDENT PERFORMANCE

Abstract

The purpose of this study is to examine the suitability of machine learning (ML) techniques for predicting students' performance. By analyzing various ML algorithms, the authors assess the accuracy and reliability of these approaches, considering factors such as data quality, feature selection, and model complexity. The findings indicate that certain ML methods are more effective for student performance forecasting, emphasizing the need for a deliberate evaluation of these factors. This study provides significant contributions to the field of education and reinforces the growing use of ML in decision-making and student performance prediction.

1. INTRODUCTION

Student performance is a crucial aspect of education, as it reflects the level of understanding and mastery of the subject matter by individual students. The assessment of student performance is essential for ensuring the quality of education and for providing feedback to students, teachers, and parents. In secondary education, student performance is particularly critical, as it can have a significant impact on future academic and career opportunities.

The traditional methods of assessing student performance, such as grades, attendance records, and standardized test scores provide valuable information about student strengths and weaknesses. However, these methods can also be subject to measurement error and biases, making it difficult to assess student performance accurately.

The increasing availability of educational data and advancements in ML techniques have led to growing interest in using these methods to improve the accuracy and reliability of student performance assessment. ML approaches have the potential to provide more accurate and reliable predictions than traditional methods, as they can automatically identify patterns and relationships in the data that may not be easily visible to the human eye.

The application of ML for forecasting student-learning outcomes offers a wealth of opportunities for educational institutions. By analyzing the learning patterns and behaviors of students, these institutions can gain a deeper understanding of their students and use this information to make informed decisions regarding school policies, curriculum development,

* Lebanese International University, Sciences, Computer, Lebanon, bilal.owaidat@liu.edu.lb

and teaching methods. ML algorithms have the potential to provide highly accurate forecasts and improve decision-making processes within educational organizations. Accurate forecasting is essential to avoid major problems and system failures, and ML can help to minimize forecasting errors. With ML models that monitor student progress and provide personalized recommendations, learning analytics become possible. This paper's goal is to explore the use of ML approaches for student performance assessment in secondary education. This study uses historical student performance data as input and applies multiple ML models to make predictions about future performance. It analyzes data and contrast it to determine the best method for assessing student performance in secondary education. The findings of this study provide valuable insights for educators and administrators, as they have the potential to inform data-driven decisions in secondary education.

2. LITERATUR REVIEW

Several academics have recognized the growing benefits and application of data mining (DM) and machine learning as a forecasting tool in the academic sector. With proper data processing and filtering, machine learning produced a variety of methods or algorithms in recent years to forecast scenarios based on vast volumes of information that can produce extremely accurate forecasts. Applying ML to education areas such as academic performance (Ahajjam et al., 2022), student perception (Demir & Güraksın, 2022) and teacher perception (Salas Rueda et al., 2022) shows the implications of using intelligent techniques in the solution of complex problems in the education sector.

The authors of (Chen & Zhai, 2023) focused on using a single type of educational data to anticipate student success. The evaluation indicates that Artificial Neural Network and Decision Tree models show promise for predicting students' performance when Random Forest is used. In (Adane et al., 2023), the authors address the challenge of predicting academic attainment effectively and reliably. The technologies employed in this work include Random Forest, Multilayer Perceptron, Naïve Bayes, and C4.5 decision Tree.

(Alghamdi & Rahman, 2023) identified the components that affect academic achievement. Naïve Bayes, Random Forest, and Synthetic Minority Oversampling Technique are used in this study. (Kukkar et al., 2023) employed Gradient Boosting, Random Forest, and Long Short Term Memory network to analyze students' performance on a variety of assessments. Even though Deep Learning technologies are superior, there is a significant difference in their predictive power.

The article by Onyema et al. (2022) examines two models for predicting student's performance in final exams based on a dataset from the University of Minho in Portugal. The dataset, consisting of 395 math performance samples, was used to explore improved prediction models. Previous studies using the K Nearest Neighbor (KNN) algorithm produced low results, so both the Support Vector Machine (SVM) and KNN algorithms were applied to the dataset to compare their accuracy. The results showed that the SVM performed better than KNN.

The article (Fayoumi & Hajjar, 2020) introduces a novel way of using Artificial Intelligence (AI) to forecast academic performance in higher education. The authors conducted research in Saudi Arabia and aimed to improve performance by incorporating data mining (DM) and decision-making techniques. They created a decision support system

based on an Artificial Neural Network Model (ANN) model, which evaluates academic metrics and predicts student performance. The accuracy of the system was tested using real data and compared to other mathematical methods. This work represents a new approach to using AI for informed decision making in higher education, with future plans to integrate with big data and analytics.

The authors of the paper (Yang et al., 2022) use AI to analyze education data to identify the factors that impacts student performance. They use DM, analysis, and visualization techniques to analyze two educational data sets. They also apply feature selection and improve the prediction of student performance using K-means and Deep Neural Network (DNN). The results indicate that the proposed Adaptive-K-means-DNN model has better performance and that the factors that affect student performance include mother's education, classroom absences, and encouragement.

The authors of the study (Agrawal & Mavani, 2015) propose a method for forecasting student performance in an academic setting using Neural Networks, an ML approach. The paper evaluates the significance of multiple attributes in determining their relationship to student performance. The outcome of the experiment highlights the usefulness of ML for predicting student performance.

The paper (Sekeroglu et al., 2019) emphasizes the significance of education for a better life and how AI is being utilized in higher education to improve the teaching and learning process. The authors experiment with two datasets to predict and classify student performance using five ML algorithms. The preliminary results indicate that preprocessing the raw data improves the accuracy of predicting and classifying student performance.

The authors of (Harvey & Kumar, 2019) explore the application of predictive classifiers for analysis of K-12 student performance data. They create and compare models using linear regression (LR), decision tree (DT), and Naive Bayes (NB) techniques. The NB approach demonstrated the greatest accuracy in predicting high school students' SAT Math scores. Stakeholders in K–12 education can utilize these insights to forecast and put strategies into place that will improve student performance quickly. The paper (Gull et al., 2020) presents a study on the use of ML to predict students' grades in an undergraduate course. The aim is to help academics optimize their teaching strategies and improve the learning experience. The study applied multiple ML techniques on historical student grades data and found that linear discriminant analysis was the most effective method, achieving an accuracy of 90.74% in predicting students' final exam performance.

The research (Altabrawee et al., 2019) aims to improve students' education at Al-Muthanna University by using ML to predict performance in a computer science course. Four methods, such as ANN, NB, DT, and LR were applied and the impact of internet use and time spent on social networks on student performance was evaluated. The models were evaluated based on their receiver operating characteristic (ROC) curve index and classification accuracy, with the ANN model achieving the highest accuracy of 77.04%. The DT model also identified the five factors that affect student performance.

The paper (Xu et al., 2017) presents a new ML technique to predict student performance in degree programs, addressing challenges such as student background diversity and varying course relevance. It features a two-layer structure with multiple base predictors and a sequence of ensemble predictors, and utilizes latent factor models and probabilistic matrix factorization to identify the significance of courses. The method was tested on data from UCLA and demonstrated better results compared to standard methods.

The work (Oyedeji et al., 2020) aimed to enhance student performance by examining past academic records and individual characteristics like age, demographics, family background, and study attitude using ML techniques. Three models were tested, namely Linear Regression for Supervised Learning, Linear Regression combined with Deep Learning, and a Neural Network. The Linear Regression for Supervised Learning model had the lowest mean average error (MAE) and produced the best results. The research (Waheed et al., 2020) looks into using a deep ANN to identify students at risk of underperforming in virtual learning environments by analyzing their learning behavior and performance data. The model was developed based on unique features extracted from virtual learning environment clickstream data. The results showed that the model had a classification accuracy of 84-93%, surpassing the performance of LR and SVM models. The study also highlights the significance of including legacy data and assessment data in the model, with students who access previous lecture content performing better. The goal of the study is to aid educational institutions in developing a framework for effective pedagogical support and data-driven decision-making in higher education.

The paper (Ghorbani & Ghousi, 2020) evaluates different resampling techniques for handling imbalanced data in student performance prediction using ML classifiers. The study analyzes the impact of class and feature structure on the prediction results. The evaluation methods used include Random holdout and Shuffle 5-fold cross-validation. The findings indicate that fewer classes and nominal features result in better performance, and balancing the data improves the performance of the classifiers. The outputs of the Friedman test show that SVM-SMOTE is the most effective resampling technique, and the Random Forest (RF) classifier produces the best results when combined with SVM- SMOTE.

3. MATERIALS AND METHODS

This section describes the approach employed for forecasting student's performances. The methodology used is detailed in Figure 1.

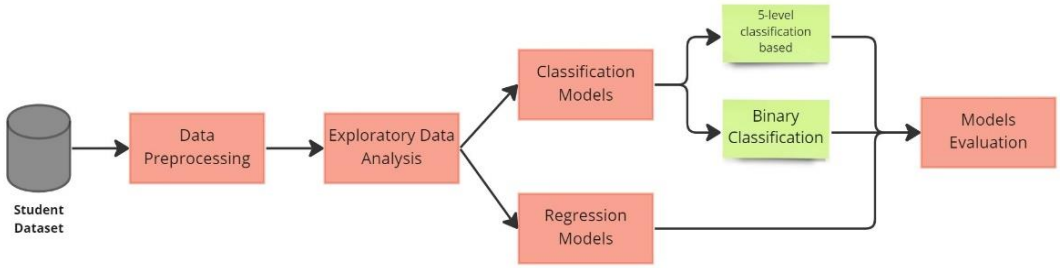

Fig. 1. Approach for forecasting students performances

3.1. Student dataset

In Portugal, secondary education is a three-year program following nine years of basic education and preceding higher education. Most students attend the public and free education system, where various courses like Sciences and Technologies and Visual Arts are offered. These courses share fundamental disciplines including Portuguese and

Mathematics. A 20-point system is used for grading, with zero being the lowest grade and 20 being the highest. Students are examined three times a year, with the most recent evaluation determining their final grade.

This study used data from the 2005-2006 school year from two public schools in the Alentejo region of Portugal. The data was collected from school reports and questionnaires, with the latter used to gather information on demographics, socio-emotional factors, and school-related variables believed to impact student performance. After preprocessing, the data was integrated into datasets for Mathematics and Portuguese language classes. Table 1 lists the remaining features, including those taken from the school reports.

Tab. 1. Student features after preprocessing

Feature	Description
Address	Address type of student's home (Urban or Rural)
Absences	Number of school absences (0 to 93)
Activities	Extra-curricular activities (Yes or No)
Age	Age of student (15 to 22)
Dalc	Weekday alcohol consumption (1 to 5)
Failures	Number of past class failures (1 to 4)
Fedu	Education of student's father (0 to 4)
Fjob	Occupation of student's father
Famsize	Family size (Less than or equal to 3 or More than 3)
Famrel	Family relationship quality (1 to 5)
Famsup	Educational support from family (Yes or No)
Free time	Free time after school (1 to 5)
G1	First period grade (0 to 20)
G2	Second period grade (0 to 20)
G3	Final grade (0 to 20)
Goout	Time spent with friends (1 to 5)
Guardian	Guardian of student (Mother, Father, or Other)
Health	Current health status (1 to 5)
Higher	Desire for higher education (Yes or No)
Internet	Internet access at home (Yes or No)
Medu	Education of student's mother (0 to 4)
Mjob	Occupation of student's mother
Nursery	Attendance at nursery school (Yes or No)
Paidclass	Paid classes outside of school (Yes or No)
Pstatus	Parental cohabitation status (Living together or Apart)
Reason	Reason for choosing school (Close to home, school reputation, course preference, or other)
Romantic	Romantic relationship (Yes or No)
School	Name of student's school (Gabriel Pereira or Mousinho da Silveira)
Schoolsup	Educational support from school (Yes or No)
Sex	Gender of student (Male or Female)
Studytime	Weekly study time (1 to 4)
Traveltime	Travel time from home to school (1 to 4)
Walc	Weekend alcohol consumption (1 to 5)

3.2. Data preprocessing

Data preprocessing is a significant step in the data analysis process, aimed at getting the raw data into a usable and meaningful format. It involves cleaning, transforming, and normalizing the data to ensure its suitability for further analysis or modeling. In this study,

the authors employed a two-step process to preprocess the data. The first step involved encoding the output variable, which is a categorical feature. To do this, the Label Encoder technique was utilized. This technique assigns numerical values to each unique category in the output variable, allowing the authors to use it as an input to their ML model.

The second step involved converting the categorical features in the input data into numerical ones. To do this, the authors used the "get_dummies" technique which creates a new binary feature for each unique category in the original feature. For example, if a feature had three categories (A, B, and C), the get_dummies technique would create three new binary features, one for each category. This process allows us to use the categorical data as input to our ML model, which typically only accepts numerical input.

By using these two techniques, the authors were able to effectively preprocess the data and prepare it for further analysis and modeling.

3.3. Exploratory data analysis

The study began by drawing several graphs on the dataset. Figure 2 shows that the majority of students in a romantic relationship perform adequately. Whereas the number of kids that fail and pass is almost the same. Students in a certain grade group who are in relationship and those who are not tend to be similar. Figure 3 depicts a violin plot, which is a graph that displays numerical data as well as the probability density distribution at various values. It shows that students above the age of 18 are more likely to pursue higher education and spend more time studying than others. Before the age of 18, the median of study time decreases as fewer pupils are interested in further education.

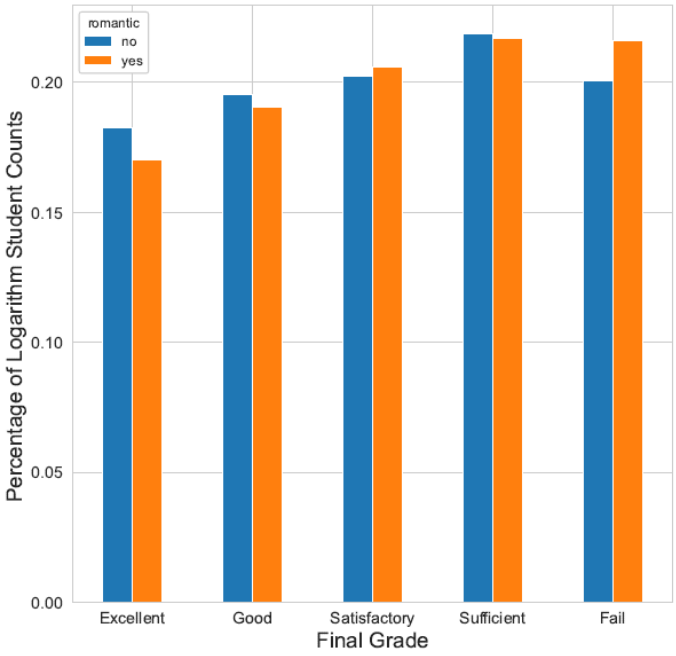

Fig. 2. Final grade by romantic status

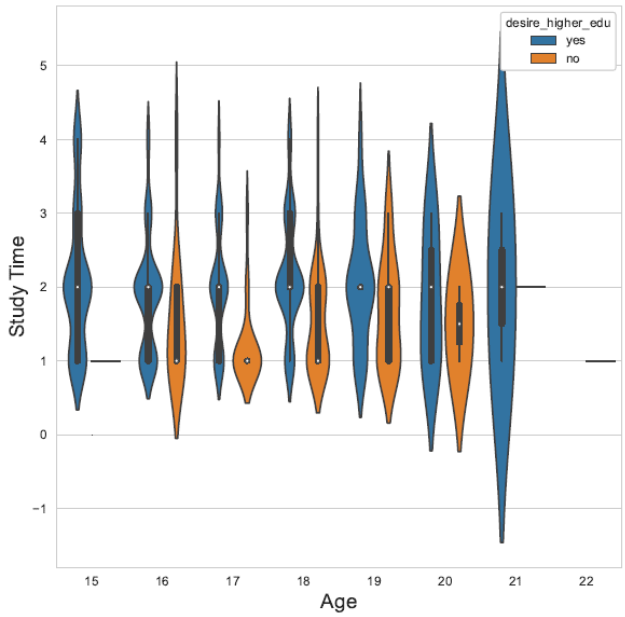

Fig. 3. Distribution of study time by age & desire to receive higher education

3.4. Model selection

DM relies heavily on classification and regression. They both need supervised learning, which involves training a model on a dataset of N samples, each of which has an input vector and a class value. The primary distinction is in the output representation, with classification giving discrete outputs and regression producing continuous outputs.

3.4.1. Classification algorithms

Classification in machine learning refers to the process of labeling given input data items into predetermined groups (William & Badholia, 2020). Many classification algorithms exist which are focused on trees, bayes, functions, or laws that are commonly used. However, there have been many inquiries about how competent these algorithmic techniques are, which made them the target of many studies. In this study, the authors target four of them (Decision Tree, Random Forest, Gradient Boosting and XGBoost).

Decision Tree is a predictive model that uses a tree-like structure to make decisions or predictions in classification tasks. DT splits the data based on the values of features, recursively creating branches and internal nodes. Each internal node represents a feature, and each branch represents a possible value or outcome of that feature. The process continues until reaching leaf nodes, which represent predicted class labels or outcomes. The decision tree algorithm learns the optimal splits from labeled data during training. It aims to maximize the separation of different classes or outcomes using criteria like information gain or Gini impurity. Decision trees are interpretable, handle various data types, and can handle multi-class classification tasks.

Random Forest is a powerful ML algorithm that can be employed for both binary and multi-class classification (Breiman, 2001). In binary classification, RF employs multiple

decision trees to predict the class membership of an observation based on its input variables. The predictions from each tree are combined through majority voting to produce the final prediction. This ensemble approach helps to reduce overfitting, increase stability and improve accuracy. In multi-class classification, the algorithm extends the binary approach by assigning each tree to predict one class and selecting the class with the most votes as the final prediction. The feature selection process in Random Forest uses a subset of the available variables for each tree, further reducing overfitting and improving accuracy. This algorithm can also handle non-linear relationships between the input variables and target class, making it a useful tool for many classification problems.

Gradient Boosting is a ML technique suitable for both binary and multi-class classification tasks (Natekin & Knoll, 2013). In binary classification, GB builds an ensemble of decision trees through a sequential process, where each new tree aims to correct the errors made by the previous trees. The final prediction is achieved by combining the predictions of all trees. This approach leads to improved accuracy. For multi-class classification, Gradient Boosting can either use the one-vs-all approach, where binary classifiers are fit for each class, or directly model the class probabilities. Gradient Boosting optimizes the predictions by adjusting the input variable weights, thereby focusing on the most relevant variables. Additionally, the technique can handle non-linear relationships, making it an effective solution for various classification problems.

XGBoost, or eXtreme Gradient Boosting, is a widely used ML algorithm for binary and multi-class classification (He, 2015). It builds an ensemble of decision trees in a sequential manner, similar to gradient boosting, but with added optimizations for improved accuracy. In binary classification, the final prediction is made by combining the predictions of all decision trees. For multi-class classification, XGBoost can either use the one-vs-all approach, where individual binary classifiers are trained for each class, or it can directly model class probabilities. XGBoost includes features such as regularization, early stopping, and parallel processing to handle high-dimensional data and prevent overfitting. Its combination of accuracy and additional features make it a popular choice for classification tasks.

3.4.2. Regression algorithms

Regression is a valuable tool for modeling complex relationships and making predictions, making it a useful technique in various industries, such as finance, economics, biology, and engineering. The aim of regression is to identify the optimal mathematical relationship between the dependent and independent variables. This relationship can be represented by either a linear or a non-linear equation, depending on the nature of the association (Freund, et al., 2006). The predicted values of the dependent variable are then determined based on the values of the independent variables. The selection of the most suitable regression model will be determined by the data characteristics and the problem being addressed. This study focuses on the following seven regression algorithms (Decision Tree, Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Gaussian Process Regression and XGBoost.

Decision tree can also be used for regression problems, where the goal is to predict a continuous target value (Kingsford & Salzberg, 2008). It builds a tree-like model by repeatedly splitting the data based on the values of the input features, selecting the feature

that minimizes the variance or mean squared error of the target values in the resulting subsets. The final prediction is made by taking the average of the target values in the terminal leaves that a data point falls into. This process results in a simple to understand and interpret model, but it can also be prone to overfitting, so techniques such as pruning or limiting the tree depth are employed to avoid this.

Linear Regression is a commonly used ML method for regression tasks that models the relationship between a target variable and one or more input features as a linear equation (Su, Yan and Tsai, 2012). The objective of Linear Regression is to find the line of best fit that minimizes the difference between the observed target values and the values predicted by the linear equation. The line of best fit is represented as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 + \cdots + \beta_p x_p \quad (1)$$

where y is the target, x_1, x_2, ..., x_p are the input features, β_0 is the y-intercept, and β_1, β_2, ..., β_p are the coefficients that indicate the relationship between the inputs and target. It's significant to keep in mind that Linear Regression assumes certain properties about the relationship between inputs and target, including linearity, independence, homoscedasticity, and normality. If these assumptions are not met, the performance of the model may be affected, and alternative algorithms should be considered. Despite these limitations, Linear Regression is a quick and interpretable algorithm that works well with both single and multiple input features. It is widely used across various fields due to its simplicity and effectiveness.

Ridge Regression's objective is to decrease the residual sum of squares (RSS) along with the L2 regularization term (McDonald, 2009), represented as:

$$J(w) = RSS + \lambda ||w||_2^2 \quad (2)$$

where w is the vector of coefficients, λ is the regularization parameter, and $\lambda ||w||_2^2$ is the L2 norm of the coefficients. The larger the value of λ, the stronger the regularization, resulting in smaller coefficients and preventing overfitting. Ridge Regression is particularly effective for datasets with high dimensionality and a risk of overfitting. The regularization helps to reduce variance and improve the generalization performance of the model. However, it may also lead to under fitting if the regularization is too strong.

Lasso Regression is a regularized linear regression algorithm that aims to address overfitting by adding a penalty term to the cost function (Ranstam and Cook, 2018). The penalty term is proportional to the absolute magnitude of the coefficients, which helps to reduce their size and prevents overfitting. The objective in Lasso Regression is to decrease the residual sum of squares (RSS) along with the L1 regularization term:

$$J(w) = RSS + \lambda ||w||_1 \quad (3)$$

where w is the vector of coefficients, λ is the regularization parameter, and $||w||_1$ is the L1 norm of the coefficients. The value of λ determines the strength of the regularization, with larger values resulting in smaller coefficients and stronger regularization. Lasso Regression is particularly useful for feature selection, as the regularization helps to identify

and eliminate redundant or irrelevant features. However, the regularization can also lead to under fitting if it is too strong.

Elastic Net Regression is a linear regression model that combines the Lasso and Ridge regression techniques (Zou & Hastie, 2005). It is used when there are more predictors than observations or when predictors are correlated. It adds a penalty term to the traditional regression model, consisting of both L1 (Lasso) and L2 (Ridge) regularization. The L1 penalty encourages sparse solutions and performs feature selection by shrinking some coefficients to zero. The L2 penalty encourages small non-zero coefficients and helps with multi-collinearity. The model has two tuning parameters ($\lambda 1$ and $\lambda 2$) that control the amount of regularization. The mixing parameter α determines the balance between L1 (Lasso) and L2 (Ridge) penalties. It ranges between 0 and 1, where 0 corresponds to Ridge regression and 1 corresponds to Lasso regression

$$J(w) = RSS + \lambda_1 * \alpha * ||w||_1 + \lambda_2 * (1 - \alpha) * ||w||_2^2 \quad (4)$$

Gaussian Process Regression (GPR) is a non-parametric ML method for regression tasks (Schulz et al., 2018). It models the relationship between the target and input features as a Gaussian distribution and makes predictions based on this distribution. In GPR, the objective is to find the Gaussian distribution that best fits the observed data. The distribution is defined by a mean function and a covariance function, which capture the relationship between the target and inputs. The mean function is typically set to zero, while the covariance function defines the relationship between each pair of points in the data. One advantage of GPR is that it provides a probabilistic prediction, which can be used to calculate the uncertainty of the prediction. This can be useful in cases where a high level of uncertainty is acceptable, such as in the exploration of new data. Another advantage of GPR is that it can handle non-linear relationships between the inputs and target and can also handle missing data. However, it can be computationally expensive, particularly for large datasets. The choice of covariance function is an important part of the GPR model and can have a considerable impact on the achievement of the model. Common covariance functions include the radial basis function (RBF) and the Matérn covariance function.

XGBoost can also be used for regression tasks. It combines weak predictive models in a boosting framework to create a robust regression model (He, 2015). XGBoost iteratively improves predictions by fitting new trees to the negative gradients of the loss function. It incorporates techniques such as regularization, tree pruning, and column subsampling to enhance performance and prevent overfitting. By tuning hyper-parameters like the learning rate, tree depth, and regularization, XGBoost optimizes its regression performance.

3.5. Model evaluation

The Percent of Correct Classification (PCC) is the classification assessment measure, while the Root Mean Squared is the regression performance measure (RMSE) (Chai & Draxler, 2014). A high PCC or a low RMSE indicates a good model. We will use the RMSE in our model.

$$\emptyset(i) = \begin{cases} 1, if\ y_i = y_{b,i} \\ 0, else \end{cases} \quad (5)$$

$$PCC = \frac{1}{N}\sum_{i=1}^{N} \emptyset(i) * 100 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - y_{b,i}\right)^2} \quad (7)$$

4. RESULTS

In this study, three supervised approaches are used to model the Mathematics and Portuguese grades (G3):

- Binary classification (pass/fail)
- 5-level classification based on a grading system
- Regression: the G3 value (numeric output between 0 and 20)

4.1. Binary classification

The target variable, final_score, is converted into a binary variable to simplify the prediction task. The final_score, which ranges between 0 and 20, is separated into two classes: pass and fail (Figure 4). The threshold for separating the classes is set to 10. The conversion of the target variable into a binary variable transforms the original problem of predicting a continuous value into a binary classification task, where the goal is to predict one of two classes: pass or fail. This transformation can sometimes result in improved model performance and more interpretable results.

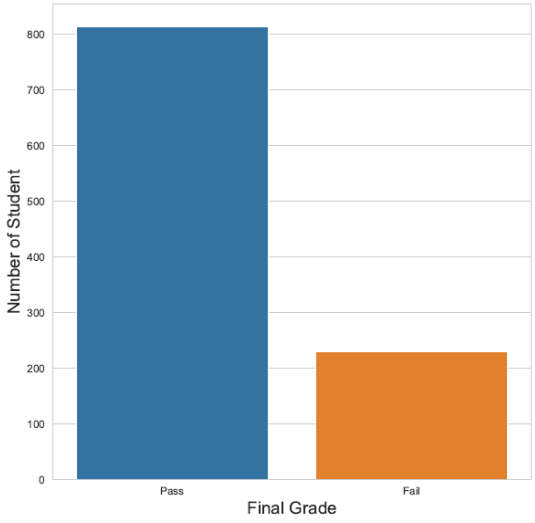

Fig. 4. Binary classification

The comparison of four binary classification models is presented in the Table 2, including DT, RF, GB, and XGBoost. The assessment is based on three key performance indicators: Accuracy, Precision, and Recall. The results reveal that RF model has the highest accuracy with score of 0.96, followed by DT and GB with 0.95 while XGBoost has a slightly lower

accuracy of 0.94. In terms of precision (PR), all four models exhibit a score of 0.92 or higher, with the RF model showing the highest precision of 0.96 and DT showing the lowest of 0.92. As for recall (RE), the RF model boasts the highest recall of 0.94, followed by GB with a recall of 0.93, and then DT and XGBoost with the lowest recall of 0.92.

Tab. 2. Accuracy of binary classification models

Models	ACC	PR	RE
Decision Tree	0.95	0.92	0.92
Random Forest	0.96	0.96	0.94
Gradient Boosting	0.95	0.95	0.93
XGBoost	0.94	0.95	0.92

The confusion matrix and the ROC curve for the binary classification using random forest are presented in Figure 5 and Figure 6 respectively.

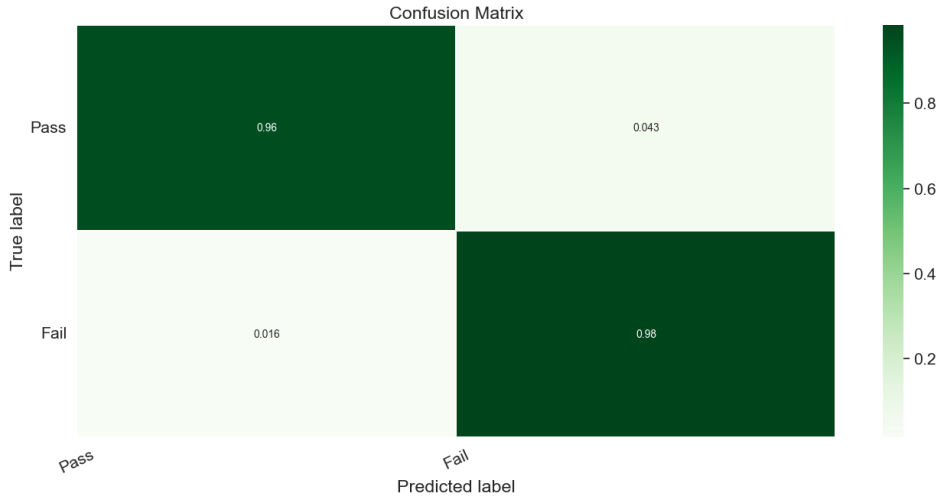

Fig. 5. Confusion matrix for binary classification using random forest

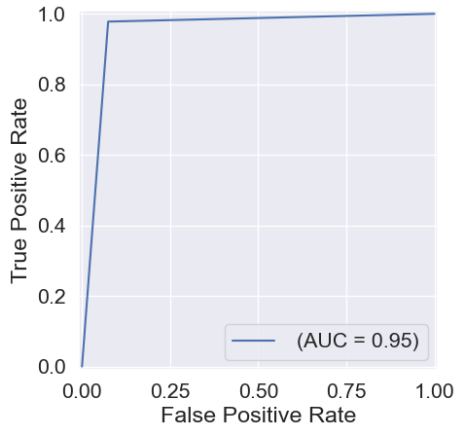

Fig. 6. ROC curve for binary classification using random forest

4.2. 5-level Classification based on a grading system

5-level classification is a machine learning task where the goal is to categorize the input data into one of five classes based on certain features or attributes. The output of a 5-level classifier is a class label from a set of five possible labels. It is a type of supervised learning, where the model is trained on labeled data to make predictions (Grandini et al., 2020). The target variable, final_score, which ranges between 0 and 20, is converted into a 5-level categorical variable (Figure 7). The goal of this conversion is to introduce a new feature into the model that can capture non-linear relationships between the target variable and the final_score. The final_score is divided into five ranges: excellent, good, satisfactory, sufficient, and fail as illustrated in Table 3. This categorization into five classes allows for a more nuanced and detailed prediction than binary classification. The new categorical variable can be used in place of the original continuous variable in any machine-learning model.

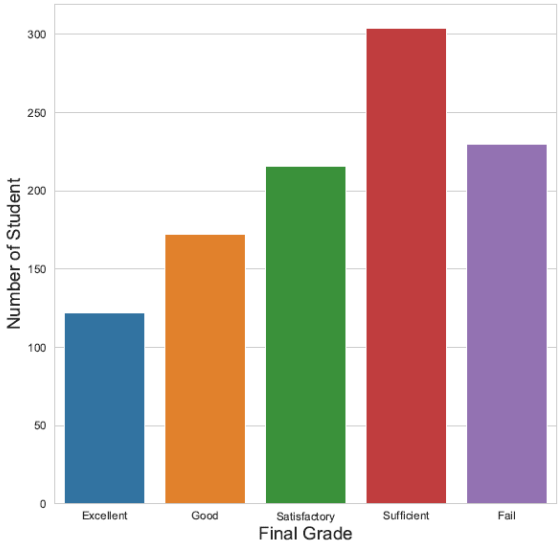

Fig. 7. Five level classification

Tab. 3. The five-level classification system

Country	(Exellent / Very Good)	(Good)	(Satisfactory)	(Sufficient)	(Fail)
Portugal / France	16-20	14-15	12-13	10-11	0-9
Ireland	A	B	C	D	F

Table 4 displays the performance comparison of our 5-level classification models, including DT, RF, GB, and XGBoost, evaluated using three metrics: Accuracy, Precision, and Recall. The RF model has the highest accuracy with 0.9, followed by GB and DT with 0.89 and then XGBoost with 0.84. The RF model has the highest precision of 0.91, followed by GB with 0.9, DT with 0.88, and XGBoost with 0.86. In terms of recall, RF showed the highest with 0.9, followed by DT and GB with 0.89 and then XGBoost having the lowest with 0.85.

Tab. 4. Accuracy of 5-level classification models

Models	ACC	PR	RE
Decision Tree	0.89	0.88	0.89
Random Forest	0.9	0.91	0.9
Gradient Boosting	0.89	0.90	0.89
XGBoost	0.84	0.86	0.85

The confusion matrix for the five level classification using random forest is presented in Figure 8.

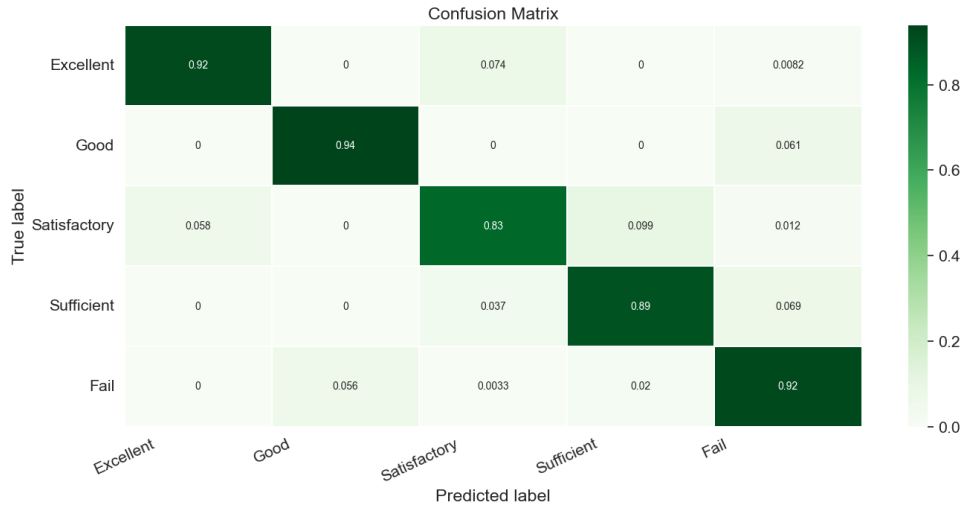

Fig. 8. Confusion matrix for the five level classification using random forest

4.3. Regression results

Table 5 displays the precision of six models utilized for 5-level classification, including Decision Tree, Linear egression, Ridge Regression, Lasso Regression, Elastic Net Regression, GPR, and XGBoost. To provide fair comparison among all models, we use the same regularization parameter (alpha = 0.05) for all scenarios. The accuracy of each model is evaluated using two metrics: MAE and RMSE. The smaller the MAE and RMSE values, the higher the model's accuracy. According to the table, Lasso Regression stands out as the most accurate model, with the smallest MAE value of 0.82 and RMSE of 1.26.

Tab. 5. MAE and RMSE of regression models

Models	MAE	RMSE
Decision Tree	0.89	1.33
Linear Regression	0.82	1.2
Ridge Regression	0.82	1.2
Lasso Regression	0.78	1.15
Elastic Net Regression	0.79	1.18
Gaussian Process Regression	0.82	1.19
XGBoost	0.8	1.17

5. DISCUSSION

Based on the results shown in the tables, the best task for data appears to be binary classification, as the models in Table 2 (Logistic Regression, Random Forest, Gradient Boosting, and XGBoost) all had high accuracy scores, with Random Forest having the highest accuracy scores of 0.96. In comparison, the models in Table 4 (Decision Tree, Random Forest, Gradient Boosting, and XGBoost) had lower accuracy scores for 5-level classification, with the highest score being 0.9 for Random Forest. Table 5 shows the root mean square error (RMSE) and mean absolute error (MAE) of regression models (Decision Tree, Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Gaussian Process Regression, and XGBoost) as evaluation metrics. Here, the Lasso Regression model had the lowest MAE and RMSE scores respectively 0.78 and 1.15, but the performance of all models were relatively similar, with Decision Tree having the highest MAE and RMSE scores respectively 0.89 and 1.33. Overall, the results suggest that binary classification is the best task for this data based on the high accuracy scores obtained by the models in Table 2.

6. CONCLUSION

In conclusion, this study thoroughly investigated the accuracy and dependability of ML techniques for forecasting student performance. By evaluating various algorithms, the authors determined that some methods are more effective than others. The results of this study have significant implications for the domain of education. The use of ML in student performance forecasting can provide valuable information to educators and administrators, improving decision-making and supporting initiatives aimed at enhancing student outcomes.

This study confirms the potential of ML to provide more precise and reliable student performance predictions, making it a valuable asset for the education sector. To summarize, this study confirms the growing trend towards using ML for student performance forecasting and highlights the significance of considering specific factors in the forecasting process. We hope these findings will drive further research and development in this area, leading to the creation of more effective tools for student performance prediction and decision-making in education. By doing so, we can work towards improving educational outcomes for students and shaping the future of education.

7. CHALLENGES AND LIMITATIONS

Predicting student performance presents several challenges and limitations. The availability and quality of data can be a significant hurdle, as obtaining comprehensive and reliable datasets from educational institutions may be difficult. Variable selection and measurement pose additional challenges, as determining which factors are relevant and accurately quantifying them can be subjective and prone to error. Predictive models may introduce biases and perpetuate inequalities if not appropriately addressed, potentially affecting the fairness and equity of predictions. The interpretability and transparency of complex models may also be limited, making it challenging to explain the reasoning behind predictions. Additionally, the generalizability of findings across different educational

contexts may be limited, as factors such as teaching methods, student demographics, and institutional policies can vary. Lastly, the dynamic nature of student performance and ethical considerations regarding privacy and unintended consequences further add to the complexity of predicting student performance. Recognizing and addressing these challenges and limitations is crucial for advancing the field and improving the accuracy and ethical soundness of predictive models in education.

8. FUTURE WORK

Based on this study's results, more research on the following topics is advised.

- Extend the study to examine the long-term prediction of student performance. Instead of predicting short-term outcomes, we can explore how well the models can forecast academic achievement over multiple semesters or years.
- Extend the study to predict the performance of university students and employ more attributes such as work status, marital status.
- Validate the predictive models in different educational contexts or institutions. Assess the generalizability of the models by applying them to datasets from different schools, universities, or even different countries. This step would help determine the robustness and transferability of the predictive models.
- Study the effect of combining multiple algorithms together on the prediction's performance.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in repository "student" at *https://drive.google.com/drive/folders/1RNREJbgOqNOmIQayqA-0krjWOcYhk-J2.*

REFERENCES

Adane, M. D., Deku, J. K., & Asare, E. K. (2023). Performance analysis of Machine Learning algorithms in prediction of student academic performance. *Journal of Advances in Mathematics and Computer Science*, *38*(5), 74-86. https://doi.org/10.9734/jamcs/2023/v38i51762

Agrawal, H., & Mavani, H. (2015). Student performance prediction using machine learning. *International Journal of Engineering Research and Technology*, *4*(3), 111–113. http://dx.doi.org/10.17577/IJERTV4IS030127

Ahajjam, T., Moutaib, M., Aissa, H., Azrour, M., Farhaoui, Y., & Fattah, M. (2022). Predicting students' final performance using Artificial Neural Networks. *Big Data Mining and Analytics*, *5*(4), 294-301. https://doi.org/10.26599/BDMA.2021.9020030

Alghamdi, A. S., & Rahman, A. (2023). Data mining approach to predict success of secondary school students: A Saudi Arabian case study. *Education Sciences*, *13*(3), 293. https://doi.org/10.3390/educsci13030293

Altabrawee, H., Ali, O., & Qaisar, A. (2019). Predicting students' performance using machine learning techniques. *Journal of University of Babylon for Pure and Applied Sciences*, *27*(1), 194-205. https://doi.org/10.29196/jubpas.v27i1.2108

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32. https://doi.org/10.1023/A:1010933404324

Chai, T., & Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific model development*, *7*(3), 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014

He, T. (2015). *Xgboost: Extreme gradient boosting*. Data Camp. https://rdocumentation.org/packages/xgboost/versions/0.4-2

Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, *28*, 12039-12057. https://doi.org/10.1007/s10639-023-11672-1

Demir, K., & Güraksın, G. E. (2022). Determining middle school students' perceptions of the concept of artificial intelligence: A metaphor analysis. *Participatory Educational Research*, *9*(2), 297-312. https://doi.org/10.17275/per.22.41.9.2

Fayoumi, A. G., & Hajjar, A. F. (2020). Advanced learning analytics in academic education: Academic performance forecasting based on an artificial neural network. *International Journal on Semantic Web and Information Systems*, *16*(3), 70-87. https://doi.org/10.4018/IJSWIS.2020070105

Freund, R. J., Wilson, W. J., & Sa, P. (2006). *Regression analysis*. Elsevier.

Ghorbani, R., & Ghousi, R. (2020). Comparing different resampling methods in predicting students' performance using machine learning techniques. *IEEE Access*, *8*, 67899–67911. https://doi.org/10.1109/ACCESS.2020.2986809

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classifi- cation: an overview. *ArXiv, abs/2008.05756*. https://doi.org/10.48550/arXiv.2008.05756

Gull, H., Saqib, M., Iqbal, S. Z., & Saeed, S. (2020). Improving learning experience of students by early prediction of student performance using machine learning. *2020 IEEE International Conference for Innovation in Technology (INOCON)* (pp. 1-4). IEEE. https://doi.org/10.1109/INOCON50539.2020.9298266

Harvey, J. L., & Kumar, S. A. P. (2019). A practical model for educators to predict student performance in k-12 education using machine learning, *2019 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 3004-3011). IEEE. https://doi.org/10.1109/SSCI44817.2019.9003147

Kingsford, C., & Salzberg, S. (2008). What are decision trees? *Nature Biotechnology*, *26*, 1011-1013. https://doi.org/10.1038/nbt0908-1011

Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2023). Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms. *Education and Information Technologies*, *28*, 9655-9684. https://doi.org/10.1007/s10639-022-11573-9

McDonald, G. C., (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 93-100. https://doi.org/10.1002/wics.14

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, *7*(21). https://doi.org/10.3389/fnbot.2013.00021

Onyema, E. M., Almuzaini, K. K., Onu, F. U., Verma, D., Gregory, U. S., Puttaramaiah, M., & Afriyie, R. K. (2022). Prospects and challenges of using machine learning for academic forecasting. *Computational Intelligence and Neuroscience*, *2022*(1), 5624475. https://doi.org/10.1155/2022/5624475

Oyedeji, A. O., Salami Olaolu, A. M., & Abolade, F. O. R. (2020). Analysis and prediction of student academic performance using machine learning. *Journal of Information Technology and Computer Engineering*, *4*(1), 10–15. https://doi.org/10.25077/jitce.4.01.10-15.2020

Ranstam, J., Cook, J. A. (2018). Lasso regression. *British Journal of Surgery*, *105*(10), 1348. https://doi.org/10.1002/bjs.10895

Salas Rueda, R. A., De la cruz Martínez, G., Eslava Cervantes, A. L., Castañeda Martínez, R., & Ramírez Ortega, J. (2022). Teachers' opinion about collaborative virtual walls and massive open online course during the COVID-19 pandemic. *Online Journal of Communication and Media Technologies*, *12*(1), e202202. https://doi.org/10.30935/ojcmt/11305

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on gaussian pro- cess regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1–16. https://doi.org/10.1016/j.jmp.2018.03.001

Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019). Student performance prediction and classification using machine learning algorithms. *8th International Conference on Educational and Information Technology* (pp. 7-11). Association for Computing Machinery. https://doi.org/10.1145/3318396.3318419

Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(3), 275-294. https://doi.org/10.1002/wics.1198

Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from vle big data using deep learning models. *Computers in Human Behavior*, *104*, 106189. https://doi.org/10.1016/j.chb.2019.106189

William, P., & Badholia, A. (2020). Evaluating efficacy of classification algorithms on personality prediction dataset. *Elementary Education Online*, *19*(4), 3400-3413.

Xu, J., Moon, K. H., & Van Der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, *11*(5), 742–753. https://doi.org/10.1109/JSTSP.2017.2692560

Yang, X., Zhang, H., Chen, R., Li, S., Zhang, N., Wang, B., & Wang, X. (2022). Research on forecasting of student grade based on adaptive k-means and deep neural network. *Wireless Communications and Mobile Computing*, *2022*(1), 5454158. https://doi.org/10.1155/2022/5454158open_in_new

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x