International Journal of Innovative Information Systems & Technology Research 14(3):363-371, July-Sept, 2026

© SEAHI PUBLICATIONS, 2026 www.seahipublications.org ISSN: 2467-8562

DOI:10.5281/zenodo.22806934

Student Academic Performance Prediction Using Study Behaviour and Attendance-Related Features

[1] I. Igwe, *[2] E. O. Bennett *[3] N. D. Nwiabu

[1,2&3]Department of Computer Science, Rivers State University, Port-Harcourt, Nigeria
Email: [1] igwe.ibuchi@rsu.edu.ng *[2] bennett.okoni@ust.edu.ng

ABSTRACT

Student academic performance prediction is used to identify learners who require academic assistance before final results are released, but prediction models overstate their early warning value when prior grades are mixed with behaviour and attendance predictors. This study developed and evaluated a supervised machine learning model-comparison procedure using the UCI Student Performance dataset. The final grade G3 was converted into a binary pass/fail target, where G3 >= 10 represented pass and G3 < 10 represented fail or academic risk. Three feature-set conditions were examined: FS1 used study behaviour and attendance-related variables, FS2 added demographic, family, school, lifestyle, and wellbeing variables while excluding prior grades, and FS3 added prior-grade variables G1 and G2. Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbours, and Gradient Boosting were implemented in Python 3.10 with Pandas, NumPy, scikit-learn, Matplotlib, Seaborn, and Jupyter Notebook. Mathematics and Portuguese subject files were evaluated separately using stratified 60:20:20 training, validation, and test partitions, validation-based threshold selection, and test-set reporting through accuracy, precision, recall, F1-score, risk-class metrics, and confusion matrices. FS3 gave the highest overall scores in both subjects. Mathematics recorded its highest overall result with Random Forest under FS3, obtaining accuracy of 0.873, F1-score of 0.900, and risk recall of 0.923. Portuguese recorded its highest overall result with Decision Tree under FS3, obtaining accuracy of 0.908 and F1-score of 0.946, while Gradient Boosting produced the highest Portuguese risk-sensitive FS3 result with risk recall of 0.800. FS1 results show that behaviour and attendance variables contain predictive information, but the FS2-to-FS3 comparison shows that prior grades raised final performance prediction scores across the tested models. The study contributes a reproducible feature-set comparison procedure for distinguishing early behaviour-attendance prediction from prior-grade-driven prediction.

Keywords: student performance prediction, educational data mining, machine learning, study behaviour, prior grades

1. INTRODUCTION

Student academic performance is a measurable indication of how well learners meet expected educational outcomes. It is represented through examination marks, continuous assessment scores, grade point average, course completion, or final subject grades. In educational data mining, such outcomes are modelled from structured student records for academic monitoring and intervention planning [1], [2].

Machine learning is widely used for academic performance prediction because schools and learning platforms generate records that describe student background, attendance, assessment history, study behaviour, and engagement. Predictive learning analytics uses these records to estimate future academic outcomes and to identify students who require academic assistance before adverse final results occur [3]. Late identification of academic difficulty reduces the time available for remedial teaching, advising,

attendance monitoring, and study-skills support.

One difficulty is that student performance is influenced by several related variables rather than by one factor alone. Prior grades often predict final grades well because they are direct measures of previous academic achievement. When prior grades are mixed with study behaviour and attendance predictors without separate analysis, the resulting model may appear accurate while giving limited evidence about whether modifiable or semi-modifiable learning indicators are suitable for earlier intervention. The UCI Student Performance dataset notes that final grade prediction is strongly associated with first- and second-period grades and that prediction without those grades is harder but more relevant for early warning purposes [1].

Recent studies have used classical machine learning, ensemble learning, deep learning, learning analytics, web-based prediction tools, explainability, and fairness-aware evaluation for student performance prediction [4]-[6]. These studies report feasible academic prediction, but many use broad feature sets, complex model designs, or prior academic variables without isolating the contribution of study behaviour and attendance-related predictors. This leaves an evaluation gap: the behaviour-attendance feature set needs separate testing, and the effect of adding prior grades needs measurement.

The aim of this study was to develop a model that predicts student academic performance using study behaviour and attendance-related features. The article makes three contributions. First, it separates study behaviour and attendance-related variables from wider background variables and prior-grade variables. Second, it evaluates six supervised machine learning models under three controlled feature-set conditions on the Mathematics and Portuguese files of the UCI Student Performance dataset. Third, it reports both overall performance and fail/risk-class performance so that model performance is interpreted in relation to academic support rather than only pass-class prediction.

2. Related Works

Educational data mining provides the basis for extracting predictive patterns from student records. Yagci [2] compared machine learning algorithms for educational data mining and reported that student performance was predictable from available academic variables. The study is relevant to machine learning-based academic prediction, but its feature design did not centre the separate effect of study behaviour and attendance-related variables.

Predictive learning analytics research covers several educational contexts. Sghir et al. [3] reviewed a decade of predictive learning analytics and reported increasing use of classification, regression, and learning behaviour data in academic prediction. Alalawi et al. [4] also reviewed machine learning for student performance prediction and noted that model performance depends on dataset structure, preprocessing, feature engineering, and evaluation design. These reviews describe the field but do not test the feature-set comparison problem on a specific dataset.

Student prediction requires attention to interpretability and fairness because model outputs influence advising and support decisions. Kesgin et al. [5] evaluated student performance prediction with explainability and fairness analysis using the UCI Student Performance dataset. Their results showed higher performance when G1 and G2 were included, and lower performance when those grades were excluded. This finding is relevant because it confirms that prior grades dominate prediction in that dataset. Ahmed et al. [6] also emphasised explainability for educational decision-making, arguing that accuracy alone is not sufficient when model outputs affect students.

Kesgin et al. [5] is the closest benchmark study for the present work because it used the same UCI Student Performance dataset and evaluated the effect of prior-grade exclusion. Their reported results showed that Logistic Regression achieved accuracy of 0.962, F1-score of 0.962, and AUC-ROC of 0.987 when G1 and G2 were included, while the best result after excluding those grades was Random Forest with accuracy of 0.789, F1-score of 0.876, and AUC-ROC of 0.665. The present study differs by treating FS1, FS2, and FS3 as the main experimental conditions, so that study behaviour and attendance-related variables are tested separately from broader background predictors and prior-grade variables.

Behavioural learning data predict performance where detailed online traces are available. Qiu et al. [7]

used learning process and behaviour data for e-learning performance prediction and reported accuracy between 95.44% and 97.40%. Ben Said et al. [8] applied machine learning to early prediction of student academic performance in a university setting. These studies are relevant to early academic prediction, but their contexts and available data differ from structured secondary-school records containing simpler variables such as study time, support, internet access, previous failures, and absences.

Ensemble and optimisation-based methods are also used in academic prediction. Mahawar and Rattan [9] used an ensemble-oriented approach for early student academic performance prediction, while Lagares Rodriguez et al. [10] compared optimised deep learning ensembles with feature-selection-based models and reported average accuracy of 0.9548, precision of 0.9597, recall of 0.9492, and F1-score of 0.9522. These methods improve benchmark scores in some settings, but they reduce transparency and are less suitable for a small tabular experiment whose goal is to measure feature contribution. Tang and Sian [11] highlighted data preparation and feature selection, reinforcing the need to examine how feature design affects predictive performance.

The reviewed studies establish that machine learning predicts academic outcomes, but they also leave two unresolved issues. First, behaviour and attendance-related variables are often combined with wider feature sets, making their separate predictive value unclear. Second, prior-grade variables raise performance and can inflate the apparent early warning value of a model. This study responds by comparing FS1, FS2, and FS3 under the same modelling and evaluation procedure.

3. METHODOLOGY

The study used a quantitative experimental machine learning methodology. A labelled public dataset was used because the final academic outcome was already known. The experiment compared both algorithms and feature sets so that the effect of study behaviour, attendance-related variables, background variables, and prior grades could be interpreted separately.

3.1 Dataset and Target Construction

The UCI Student Performance dataset was used as the data source [1]. It contains student achievement records from two Portuguese secondary schools and provides two subject files: student-mat.csv for Mathematics and student-por.csv for Portuguese. The Mathematics file contained 395 records, while the Portuguese file contained 649 records. Each file contained 33 columns, with no missing values and no duplicate records in the executed experiment.

The original final grade G3 ranges from 0 to 20. It was transformed into a binary academic performance target:

$$y = \begin{cases} 1, & \text{if } G3 \geq 10 \\ 0, & \text{if } G3 < 10 \end{cases} \quad (1)$$

Where class 1 represents pass and class 0 represents fail or academic risk. This target design converts the grade prediction task into a classification problem. After the target vector was created, G3 was removed from the input feature matrix to prevent direct target leakage.

3.2 Prediction Pipeline Architecture

The prediction pipeline follows a supervised machine learning structure, as shown in Figure 1. The input layer receives the UCI Student Performance dataset. The target-construction stage converts G3 into a pass/fail class, after which preprocessing applies categorical encoding and numerical scaling where required. The feature-set construction stage then creates FS1, FS2, and FS3 for separate evaluation of behaviour-attendance predictors, broader background predictors, and prior-grade predictors. The model-training stage fits the six candidate machine learning models, and the evaluation stage compares their outputs using overall and risk-class metrics.

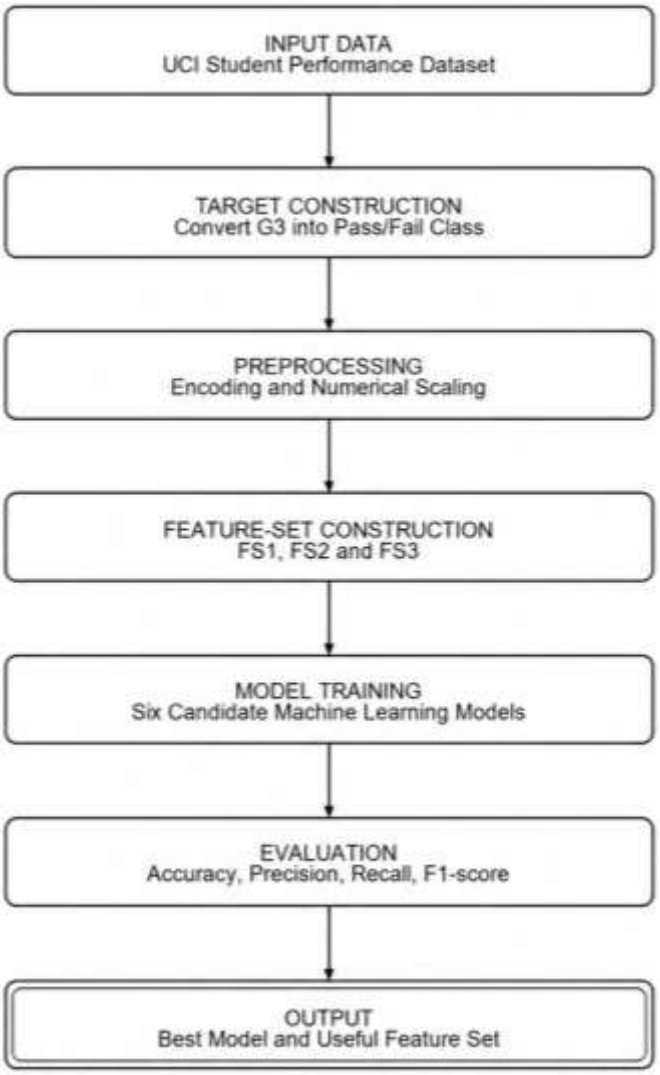

Figure 1: Prediction pipeline architecture for student academic performance analysis.

The architecture differs from a single broad-feature prediction model because feature-set comparison is built into the procedure. Each feature-set condition passes through the same preprocessing, splitting, training, prediction, and evaluation stages. This keeps the modelling procedure constant while changing the predictor group under evaluation.

3.3 Feature-Set Design

The predictor variables were separated into three feature-set conditions. FS1 contained nine behaviour and attendance-related variables: studytime, failures, schoolsup, famsup, paid, activities, higher, internet, and absences. FS2 contained 30 variables by adding demographic, family, school, lifestyle, and wellbeing predictors while still excluding prior grades. FS3 contained 32 variables by adding G1 and G2 to the FS2 predictors. FS1 therefore tested behaviour and attendance predictors alone, FS2 tested wider context without prior grades, and FS3 tested the effect of adding prior academic performance variables. ### 3.4 Model Implementation

Six supervised machine learning models were implemented: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbours, and Gradient Boosting. The implementation used Python 3.10.0, Pandas 2.3.3, NumPy 1.26.4, scikit-learn 1.7.2, Matplotlib 3.10.8, Seaborn 0.13.2, joblib 1.5.3, and Jupyter Notebook.

Categorical variables were encoded with one-hot encoding. Numerical variables were scaled where required by the algorithm. A fixed random seed of 42 was used. The data were split into 60% training, 20% validation, and 20% test partitions using stratified splitting. The validation set was used for decision-threshold selection, while the test set was reserved for final metric reporting.

Logistic Regression used max_iter=1000, balanced class weights, and random_state=42; Decision Tree used balanced class weights and random_state=42; Random Forest used 200 trees, balanced class weights, and random_state=42; Support Vector Machine used an RBF kernel, balanced class weights, probability output, and random_state=42; K-Nearest Neighbours used n_neighbors=5; and Gradient Boosting used balanced sample weights during fitting with random_state=42.

3.4 Evaluation Metrics

Model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrices, risk precision, risk recall, and risk F1-score. These metrics were computed from true-positive, true-negative, false-positive, and false-negative counts on the held-out test set. Risk-class metrics were reported because the model is intended to identify learners who require academic support, not only students likely to pass.

4. Results and Discussion

The main result was the difference between FS2 and FS3: every tested model recorded a higher F1-score after G1 and G2 were added. Both subject files were pass-class dominant: Mathematics had 395 records with 130 fail/risk cases (risk rate 0.329), while Portuguese had 649 records with 100 fail/risk cases (risk rate 0.154). This distribution makes risk recall and risk F1-score necessary because acceptable overall performance can still hide missed fail/risk students.

4.1 Overall Model Performance

Table 1 reports the highest-scoring overall model for each subject and feature-set condition. FS3 produced the highest overall scores in both subjects. In Mathematics, Random Forest under FS3 achieved accuracy of 0.873, precision of 0.957, recall of 0.849, F1-score of 0.900, and risk recall of 0.923. In Portuguese, Decision Tree under FS3 achieved accuracy of 0.908, precision of 0.938, recall of 0.955, F1-score of 0.946, and risk recall of 0.650.

Table 1: Highest-scoring overall model per subject and feature set.

Subject	Feature set	Model	Threshold	Accuracy	Precision	Recall	F1-score	Risk recall	Risk F1-score
Mathematics	FS1	Support Vector Machine	0.70	0.633	0.731	0.717	0.724	0.462	0.453
Mathematics	FS2	Decision Tree	0.10	0.633	0.722	0.736	0.729	0.423	0.431
Mathematics	FS3	Random Forest	0.52	0.873	0.957	0.849	0.900	0.923	0.828
Portuguese	FS1	K-Nearest Neighbours	0.61	0.754	0.868	0.836	0.852	0.300	0.273
Portuguese	FS2	Support Vector Machine	0.45	0.800	0.856	0.918	0.886	0.150	0.188
Portuguese	FS3	Decision Tree	0.10	0.908	0.938	0.955	0.946	0.650	0.684

The FS1 results indicate predictive value in behaviour and attendance-related variables, but their standalone performance is lower than the feature set that includes prior grades. FS1 achieved F1-score of 0.724 in Mathematics and 0.852 in Portuguese for the highest-scoring overall models. The same FS1 models had risk recall of 0.462 and 0.300 respectively, showing weaker fail/risk detection. Figure 2

shows the pass-class F1-score pattern across subjects, models, and feature sets.

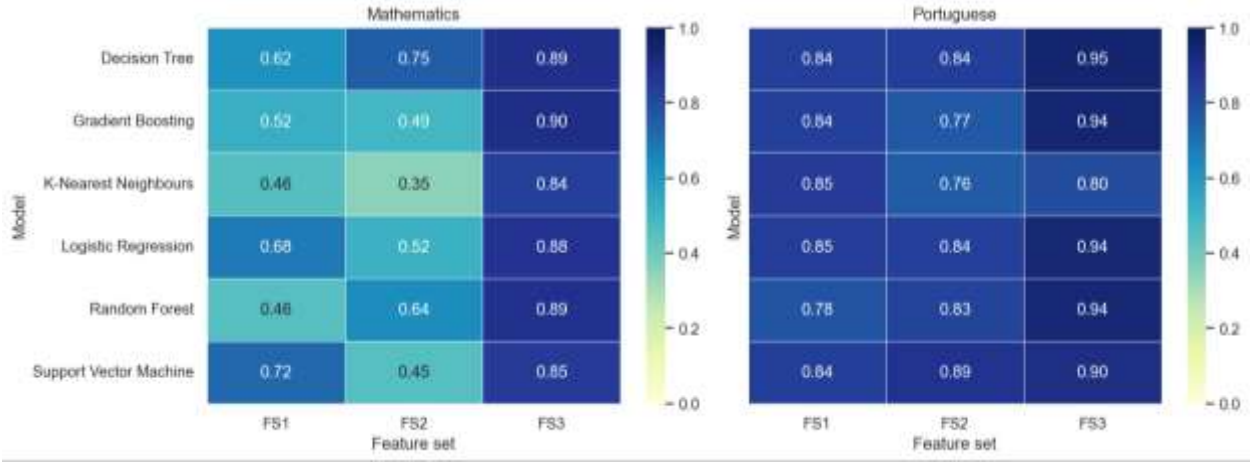

Figure 2: Pass-class F1-score heatmaps across model and feature-set conditions.

4.2 Fail/Risk-Class Detection

Table 2 reports the highest risk-sensitive result for each subject and feature set. Mathematics FS1 K-Nearest Neighbours achieved risk recall of 0.846, but with risk precision of 0.379. This means it identified 22 of 26 fail/risk students but produced 36 false risk alerts. Mathematics FS3 Random Forest gave the most balanced risk-class result in this comparison, with risk precision of 0.750, risk recall of 0.923, and risk F1-score of 0.828.

Table 2: Highest risk-class result per subject and feature set.

Subject	Feature set	Best risk-sensitive model	Threshold	Risk precision	Risk recall	Risk F1-score	Risk support
Mathematics	FS1	K-Nearest Neighbours	0.81	0.379	0.846	0.524	26
Mathematics	FS2	Random Forest	0.69	0.485	0.615	0.542	26
Mathematics	FS3	Random Forest	0.52	0.750	0.923	0.828	26
Portuguese	FS1	Logistic Regression	0.52	0.300	0.450	0.360	20
Portuguese	FS2	Random Forest	0.81	0.238	0.500	0.323	20
Portuguese	FS3	Gradient Boosting	0.60	0.667	0.800	0.727	20

The risk-class results show why overall F1-score should not be the only evidence in academic support applications. Portuguese FS2 Support Vector Machine achieved overall F1-score of 0.886, but its risk recall was only 0.150. A model with this pattern classifies many passing students correctly while missing most students who need intervention. Figure 3 shows the fail/risk recall values across the model-feature-set combinations. Confusion-matrix validation supported this interpretation: for the best overall FS3 models, Mathematics Random Forest correctly identified 24 of 26 fail/risk cases, while Portuguese Decision Tree correctly identified 13 of 20 fail/risk cases.

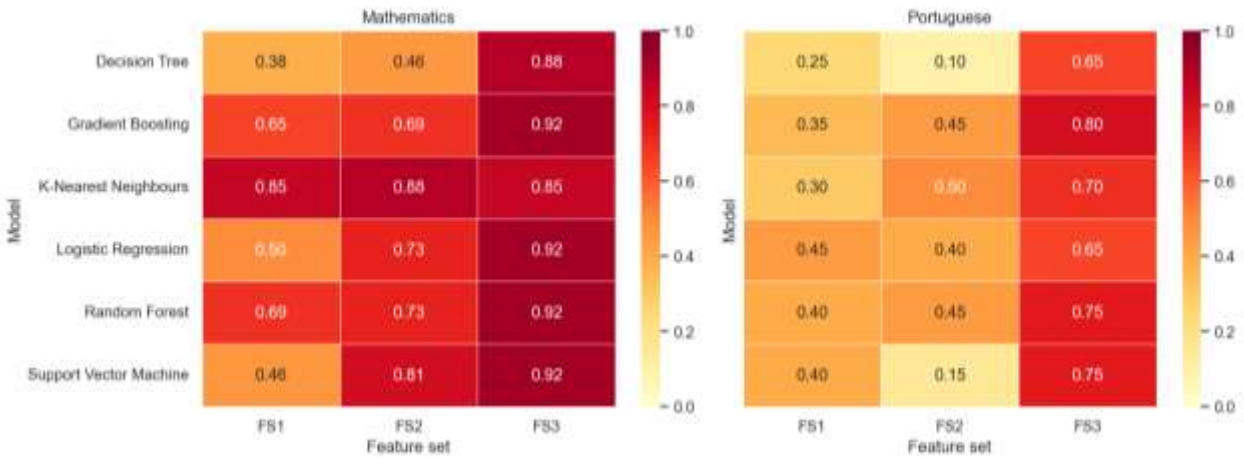

Figure 3: Fail/risk recall heatmaps across model and feature-set conditions.

4.3 Prior-Grade Effect and Feature Importance

The comparison between FS2 and FS3 isolates the effect of adding G1 and G2. All models in both subject files recorded higher F1-scores after the prior-grade variables were included. In Mathematics, the largest F1-score gains occurred for K-Nearest Neighbours (+0.484), Gradient Boosting (+0.406), Support Vector Machine (+0.401), and Logistic Regression (+0.359). In Portuguese, the largest gains occurred for Gradient Boosting (+0.158), Random Forest (+0.143), Decision Tree (+0.104), and Logistic Regression (+0.103). FS3 was therefore the highest-scoring final-performance predictor, but it depends on previous assessment results; FS1 and FS2 remain more relevant when the goal is earlier intervention before G1 and G2 are available.

Tree-based feature importance followed the same pattern: G2 was the highest-ranked predictor, followed by absences, weekend alcohol consumption, G1, health, and previous failures, as shown in Figure 4. Prior grades had the largest influence on final performance prediction, while the ranking of absences and failures shows that behaviour and attendance-related variables should remain in early warning experiments.

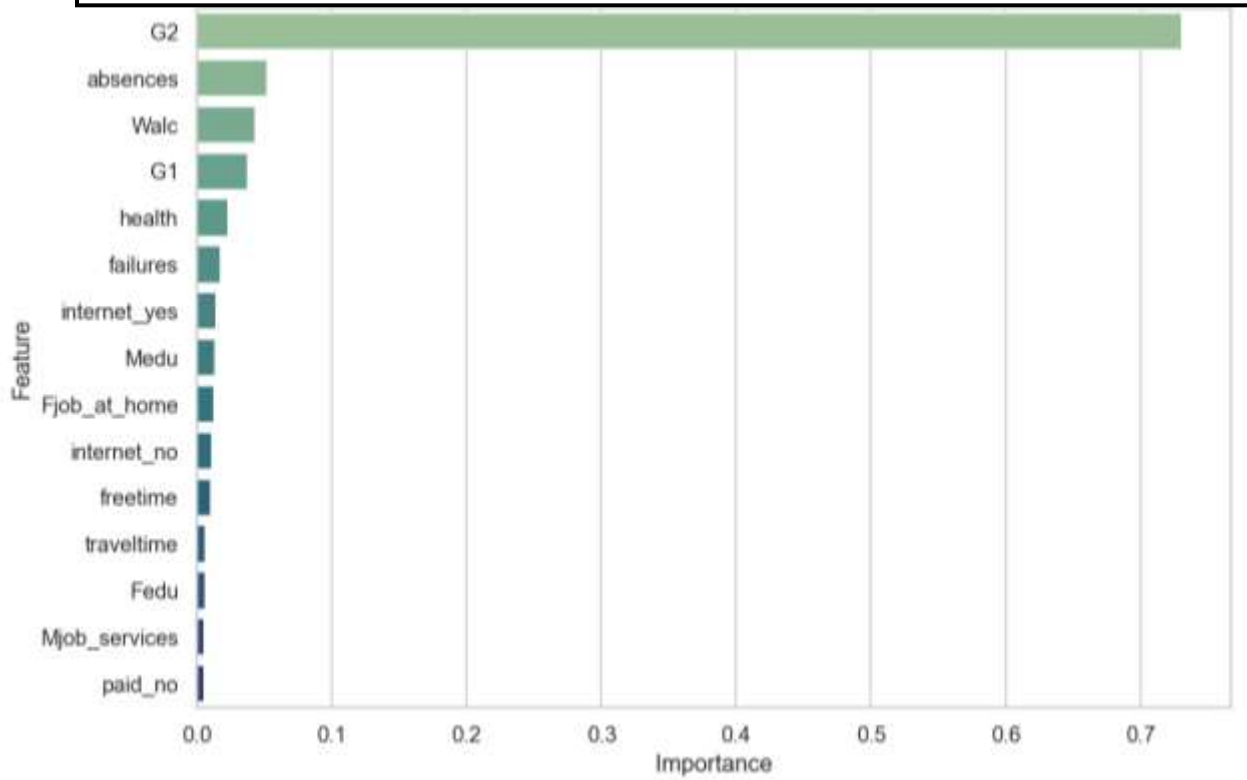

Figure 4: Feature importance for the best tree-based model.

4.4 Limitations and Benchmark Comparison

The study used a public structured dataset rather than current institutional records, and the binary pass/fail target reduced the detail in the original 0-20 final grade scale. The Portuguese file also had a smaller fail/risk proportion, making risk-class estimates more sensitive to individual classification errors. Kesgin et al. [5] provides the closest benchmark: it reported accuracy of 0.962, F1-score of 0.962, and AUC-ROC of 0.987 with G1 and G2, while the best model in this study recorded accuracy of 0.908 and F1-score of 0.946 for Portuguese Decision Tree under FS3. The article therefore does not claim a higher benchmark score; its contribution is the controlled separation of behaviour-attendance predictors, wider background predictors, and prior-grade predictors, with explicit risk-class reporting.

5. Conclusion

This article presented a supervised machine learning procedure for student academic performance prediction using study behaviour, attendance-related variables, background variables, and prior-grade feature sets. The procedure was evaluated on the Mathematics and Portuguese files of the UCI Student Performance dataset using six classical machine learning models and three feature-set conditions. FS3, which included G1 and G2, produced the highest overall scores in both subjects: Mathematics recorded F1-score of 0.900 and risk recall of 0.923 with Random Forest, while Portuguese recorded F1-score of 0.946 with Decision Tree and risk recall of 0.800 with Gradient Boosting.

Behaviour and attendance-related variables retained predictive value, but their standalone performance was weaker than the prior-grade condition. FS1 and FS2 are therefore more relevant for earlier intervention before prior assessment grades are available, while FS3 is more accurate for final performance prediction after G1 and G2 are known. The study contributes a reproducible feature-set comparison approach for distinguishing early behaviour-attendance prediction from prior-grade-driven prediction; future work should evaluate the procedure on larger institutional datasets with richer behavioural and real-time attendance variables.

REFERENCES

[1] P. Cortez, "Student Performance," UCI Machine Learning Repository, (2008). doi: 10.24432/C5TG7T.

[2] M. Yagci (2022), "Educational data mining: Prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, 11.

[3] N. Sghir, A. Adadi, and M. Lahmer (2022), "Recent advances in predictive learning analytics: A decade systematic review (2012-2022)," *Education and Information Technologies*, vol. 28, no. 7, pp. 8299-8333.

[4] K. Alalawi, R. Athauda, and R. Chiong (2023), "Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review," *Engineering Reports*, vol. 5, no. 12, e12699.

[5] K. Kesgin, S. Kiraz, S. Kosunalp, and B. Stoycheva (2025), "Beyond performance: Explaining and ensuring fairness in student academic performance prediction with machine learning," *Applied Sciences*, vol. 15, no. 15, 8409.

[6] W. Ahmed, M. A. Wani, P. Plawiak, S. Meshoul, A. Mahmoud, and M. Hammad (2025), "Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions," *Scientific Reports*, vol. 15, no. 1, 26879.

[7] F. Qiu, G. Zhang, X. Sheng, L. Jiang, L. Zhu, Q. Xiang, B. Jiang, and P. K. Chen (2022), "Predicting students' performance in e-learning using learning process and behaviour data," *Scientific Reports*, vol. 12, no. 1, 453.

[8] M. Ben Said, Y. Hadj Kacem, A. Algarni, and A. Masmoudi (2023), "Early prediction of student academic performance based on machine learning algorithms: A case study of bachelor's degree students in KSA," *Education and Information Technologies*, vol. 29, no. 11, pp. 13247-13270.

[9] K. Mahawar and P. Rattan (2024), "Empowering education: Harnessing ensemble machine learning approach and ACO-DT classifier for early student academic performance prediction," *Education and Information Technologies*, vol. 30, no. 4, pp. 4639-4667.

[10] J. A. Lagares Rodriguez, N. Diaz-Diaz, and C. D. Barranco Gonzalez (2025), "A comparative analysis of student performance prediction: Evaluating optimized deep learning ensembles against semi-supervised feature selection-based models," *Applied Sciences*, vol. 15, no. 9, 4818.

[11] L. Tang and C. Sian (2025), "Educational data mining for student performance prediction," *Scalable Computing: Practice and Experience*, vol. 26, no. 3, pp. 1551-1558.