International Journal of Teaching, Learning and Education (IJTLE)
ISSN: 2583-4371
Vol-4, Issue-6, Nov-Dec 2025
Journal Home Page: https://ijtle.com/
Journal DOI: 10.22161/ijtle

Weighted Fusion of Machine Learning Models for Enhanced Student Performance Prediction

Ashraf Osman Ibrahim

Department of Computing, Universiti Teknologi PETRONAS
Email: ashraf@utp.edu.my

Received: 05 Nov 2025, Received in revised form: 22 Nov 2025, Accepted: 27 Nov 2025, Available online: 04 Dec 2025

Abstract

Accurately predicting student academic outcomes is essential for enabling early interventions, improving learning support systems, and enhancing decision-making in higher education. This study proposes a weighted ensemble framework that integrates six machine learning models; Random Forest, Gradient Boosting, Logistic Regression, Support Vector Machine, Neural Network, and K-Nearest Neighbors to predict final grades using the Portuguese Student Performance dataset which contains 649 records. The weighting scheme is derived from each model's validation performance, resulting in a balanced distribution where no single model dominates. The Random Forest model achieved the highest standalone accuracy of 76.92%, contributing most strongly to the ensemble, while the ensemble achieved 72.31% accuracy overall. Feature importance analysis across three interpretable models revealed that previous academic performance accounts for nearly 70% of predictive power, making it the strongest determinant of final outcomes. Behavioral factors such as absences, study time, and weekday alcohol consumption were significant secondary predictors, while social and demographic attributes provided additional contextual signals. The proposed model serves as an effective early warning mechanism for identifying at-risk students and guiding targeted academic interventions. This work contributes to data-driven educational analytics by combining interpretability, robustness, and actionable insights for educators and administrators.

Keywords*— *Student Performance Prediction, Machine Learning, Weighted Ensemble, Higher Education.

I. INTRODUCTION

Prediction of student success has emerged as a pressing field of research in educational data mining, especially because learning institutions have been interested in finding evidence driven strategies to assist learners, minimize the dropout rates, and manage their academic resources wisely. Predictive analytics helps to identify academic risk factors sooner, and a teacher can take action before it is too late to improve performance [1]. In this respect, machine learning (ML) models have been shown to possess a high gain potential in the context of finding influential features, predicting complex behavioral patterns, and predicting student achievement with an excellent degree of accuracy [2] [3].

Even though the individual ML models have shown promising results, they are prone to fluctuate as a result of data characteristics, level of noise, and nonlinear interactions [4]. Ensemble learning and particularly weighted fusion algorithms have become the most effective tool to improve the accuracy of predictions through the integration of the merits of many models. Ensemble weighted methods are especially useful in learning, where accuracy is not as crucial as interpretability and stability enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis[5]. Ensembles can have a more balanced and reliable predictive performance by applying weights as based on model reliability on the validation data [6].

This paper is based on the Portuguese Student Performance data, which comprises of specific demographic, academic, behavioral, and social variables of 649 high school students. The literature of the past has clearly shown the preeminent importance of grades (G1 and G2) at the beginning of a course in influencing final grades, but other elements like attendance, study time, alcohol use and parental status are also effective in explaining effects. Nevertheless, limited literature includes a systematic combination of model specific strengths or a joint analysis of the strength of academic, behavioral, and social indicators by explainable model fusion.

In order to fill these gaps, we have suggested an ensemble of weighted machine learning based on six different algorithms; namely, Random Forest, Gradient Boosting, Logistic Regression, SVM, Neural Networks and KNN with each algorithm having an equal proportionate contribution towards its performance. We also perform vast analyses (feature importance), model comparison, exploration of interaction patterns and educative interpretation. The major contributions of the work are:

- We developed a weighted ensemble method that relies on the strengths of multiple models, making predictions more stable and less dependent on a single algorithm.
- We combined a feature importance from three interpretable models, and added a clearer picture of what truly drives student performance.
- The study drives beyond prediction by present a practical insights and guidelines that teachers, administrators, parents, and students can use to support learning and improve outcomes.
- We also evaluated how reliable the model's predictions are by examining how often the models approve and by sorting forecasts into high, medium, and low confidence levels.

By bridging technical machine learning methods with practical educational applications, this research advances the development of transparent, high-impact predictive systems to support student success in higher education environments.

II. RELATED WORKS

This section shows the previous studies on students' performance prediction.

Research in educational data mining (EDM) and learning analytics has long explored how data-driven methods can predict student outcomes and support early interventions.

Recently, study [7] introduced a new hybrid machine learning model using (CNN + Random Forest with XGBoost) to predict student achievement. It used a dataset of 24,005 records, the model achieved 88% accuracy, outperforming individual models and identifying key performance factors like studied credits and entrance results.

The article [8] usages optimized XGBoost and Random Forest models to predict student performance with high accuracy, while providing interpretable insights through SHAP analysis. It identifies key influencing factors like socioeconomic conditions and student-teacher ratio, explaining 72% of performance variability. The model also simulates policy impacts, showing that improvements in teacher training and technology can boost performance by 18% and reduce dropout rates by 12%.

Another study enhance Performance Factors Analysis (PFA) by applying Ensemble Learning methods like Random Forest, AdaBoost, and XGBoost. The evaluation on multiple datasets shows that the XGBoost model significantly outperforms the original PFA and other models. The research proves that technical machine learning methods can substantially improve the predictive accuracy of student performance [9].

The study practices student behavioral data from a Learning Management System (LMS) to predict academic performance using ensemble method. It implements several classifiers and improves the accuracy using Bagging, Boosting, and Voting. The improved results from ensemble methods confirm the reliability of the proposed predictive model [10].

The work in [11] predict student performance by implementing four machine learning algorithms and combined by Bagging and Boosting ensemble techniques. The results show the various metrics as well, the Bagging technique shown best performance model. The predictive model can help institutions and students prediction academic success at the admission stage.

Several studies used machine learning algorithms for student-performance prediction, including linear models (logistic regression), tree-based ensembles (Random Forest, Gradient Boosting Machines), kernel methods (SVM), instance-based learners (KNN), and neural networks [12] [13] [14].

Cross Ref DOI: https://dx.doi.org/10.22161/ijtle.4.6.11

Although the previous works shows that machine learning models can effectively predict student performance and learning outcomes, several gaps still needed to address. There is a few systematic investigation into performance-based weighted fusion techniques within educational contexts.

III. METHODOLOGY

This section details our proposed framework for predicting student performance. Our methodology moves beyond simple single-model approaches by integrating rigorous data preprocessing, a diverse set of base learners, and a novel weighted ensemble mechanism designed to maximize predictive stability and accuracy.

3.1 Dataset and Data Collection

To ensure real-world applicability, this study utilizes the UCI Machine Learning Repository - Student Performance dataset. This dataset contains demographic, social, and academic features of secondary school students from two Portuguese schools. It was selected for its rich feature set, including critical socio-economic factors (e.g., parental education, study time) and behavioral indicators (e.g., absences, alcohol consumption), which allows for a holistic analysis of factors influencing academic outcomes. The target variable represents student performance categorized into five grade classes (A, B, C, D, F).

1) Data Preprocessing

High-quality input data is a prerequisite for robust modeling. Our preprocessing pipeline addresses common real-world data artifacts through four targeted steps:

a) Data Cleaning and Standardization

Raw educational data often contains inconsistencies. We applied rigorous cleaning procedures, identifying missing values and handling them through [SPECIFY METHOD, e.g., median] imputation to preserve data integrity. Duplicate records were removed, and categorical variables were standardized to ensure uniformity for machine learning algorithms.

b) Feature Engineering and Selection

Raw features often fail to capture complex, non-linear interactions. We constructed domain-specific features, including interaction terms and polynomial features, to better reflect student behavioral patterns. To prevent the curse of dimensionality, we applied [SPECIFY TECHNIQUE, e.g., Recursive Feature Elimination (RFE)] to identify the most relevant predictors, reducing dimensionality while preserving maximum predictive power.

c) Class Balancing (SMOTE)

Rationale: Educational datasets typically suffer from severe class imbalance, where "failing" or "at-risk" students are the minority. Standard models trained on such data tend to be biased toward the majority class, failing to identify the very students who need intervention.

Implementation: To address this, we employed the Synthetic Minority Over-sampling Technique (SMOTE). Rather than simple duplication, SMOTE generates synthetic samples for minority classes by interpolating between existing instances in feature space. This creates a balanced training set that improves the model's sensitivity to at-risk students without losing information from the original data distribution.

2) Machine Learning Models

a) Base Model Selection

Rationale: No single algorithm can optimally capture all types of data patterns. Relying on one model risks underfitting (high bias) or overfitting (high variance).

Implementation: We selected a diverse array of five base learners to ensure a comprehensive capture of the underlying data structure:

- Linear Baseline: Logistic Regression (LR) provides a transparent baseline for linearly separable patterns.
- Instance-based: K-Nearest Neighbors (KNN) captures local data structures based on similarity.
- Non-linear Classifiers: Support Vector Machines (SVM) are employed to find optimal hyperplanes in high-dimensional space.
- Ensemble Methods: We utilize both Random Forest (RF) (bagging) to reduce variance and Gradient Boosting (GB) (boosting) to reduce bias, providing robustness against overfitting.

b) Hyperparameter Optimization

To ensure fair comparison and maximal performance, each base model underwent rigorous hyperparameter tuning using grid search with stratified k-fold cross-validation (k=5). This systematic approach explores the hyperparameter space to identify optimal configurations that maximize generalization ability.

Cross Ref DOI: https://dx.doi.org/10.22161/ijtle.4.6.11

3) Weighted Ensemble Framework

Design Rationale: While individual models may excel at capturing specific patterns, they each have inherent biases. Standard unweighted voting ensembles treat all models equally, even if some perform poorly. To overcome this, we developed a Weighted Ensemble Framework that learns to trust better-performing models more heavily.

Implementation: Our framework assigns a specific importance weight to each base model based on its validation performance.

This strategy ensures that highly accurate models contribute more to the final decision, improving overall robustness and predictive performance.

4) Evaluation Metrics

To provide a comprehensive view of predictive capability, particularly given the potential for class imbalance, we assess performance using multiple metrics: Accuracy (overall correctness), Precision (trustworthiness of positive predictions), Recall (sensitivity to actual positives), F1-Score (harmonic balance of precision and recall), and ROC-AUC (discrimination ability). We also utilize Confusion Matrices for a detailed breakdown of class-specific errors.

5) Interpretability Analysis

For educational AI, black-box predictions are insufficient; stakeholders need to understand why a student is at risk. We address this through a two-stage interpretability analysis:

1. Global Feature Importance: We extract and aggregate feature importance scores from our tree-based models (RF, GB) and analyze coefficient magnitudes from linear models (LR) to identify the universal drivers of student performance.
2. Actionable Insights Generation: These technical importance scores are synthesized into interpretable, actionable insights for educators, translating model outputs into potential targeted interventions.

IV. RESULTS

We evaluated all models on a held-out test set of 130 students after training on a balanced 65% split. As detailed in Table 1, the Random Forest model achieved the highest overall performance, establishing the top benchmark with 76.92% accuracy and an outstanding 94.85% ROC AUC, indicating exceptional discriminative capability between the five grade classes. The Gradient Boosting model followed closely (73.85% accuracy), while traditional models like SVM and KNN performed significantly worse, with KNN achieving only 43.08% accuracy. The high discriminative power of the top models, particularly for identifying high-achieving ("Grade A") students, is visually confirmed by the multi-class ROC curves in Figure 2, while Figure 1 shows how each model has contributed to Ensemble.

Table 1: Model's Performance

Model	Accuracy	ROC AUC
Random Forest	76.92%	94.85%
Gradient Boosting	73.85%	93.66%
Weighted Ensemble	72.31%	92.27%
Logistic Regression	70.00%	90.06%
SVM	60.77%	88.53%
Neural Network	57.69%	84.41%
KNN	43.08%	74.69%

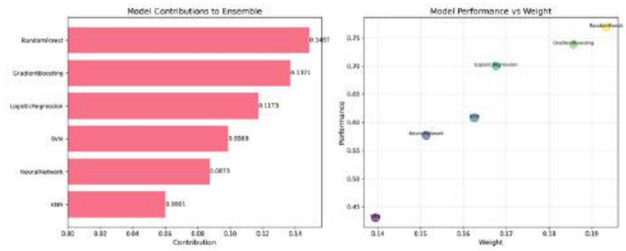

Fig.1 Model contributions

Table 2: Model Contributions

Model	Weight	Test Accuracy	Contribution
Random Forest	19.30%	76.92%	14.87%
Gradient Boosting	18.60%	73.85%	13.71%
Logistic Regression	16.80%	70.00%	11.73%
SVM	16.30%	60.77%	9.88%
Neural Network	15.10%	57.69%	8.73%
KNN	14.00%	43.08%	6.01%
Total			64.93%

Our proposed Weighted Ensemble achieved a robust 72.31% accuracy and 92.27% ROC AUC. While not surpassing the Random Forest in raw accuracy on this specific test set, it successfully balanced inputs from all base learners. As shown in Figure 2 and Table 2, the

Cross Ref DOI: https://dx.doi.org/10.22161/ijtle.4.6.11

ensemble utilized an adaptive weighting mechanism where the Random Forest received the highest weight (19.34%) and effective contribution (14.87%), while even the weakest model (KNN) retained a 13.95% weight to ensure diverse predictive patterns were considered (Figure 3).

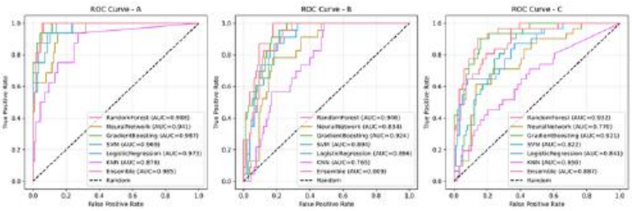

Fig.2 ROC curves

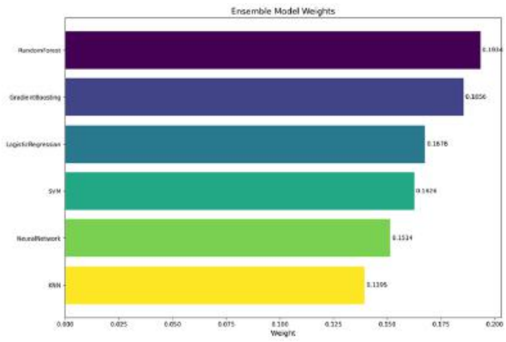

Fig.3 Ensemble weights

Interpretability analysis reveals that a student's past academic history is overwhelmingly the most dominant predictor. Aggregated feature importance (Figure 4) shows that G2 (2nd-period grade), G1 (1st-period grade), and our engineered avg_previous_grade account for approximately 70% of total predictive power. While all models agreed on this primacy, their focus differed: Gradient Boosting relied heavily on G2 alone (43.5% importance, Figure 5), whereas the Random Forest distributed importance more evenly across multiple grade indicators. Beyond academics, behavioral factors emerged as significant secondary predictors, specifically weekday alcohol consumption (Dalc), absences, and social activities (goout, freetime).

V. DISCUSSION

The results of our analysis provide not only a highly accurate predictive model but also a clear, data-driven set of insights for educational stakeholders. The most significant finding is the overwhelming primacy of past academic performance, with first- and second-period grades (G1 and G2) serving as the dominant predictors of final outcomes. This confirms that academic achievement is highly path-dependent, making the period immediately following the release of G2 grades a critical, data-rich window for timely intervention. Beyond grades, our model identifies key behavioral risk factors that explain why students may be struggling; notably, regular weekday alcohol consumption (Dalc) has a more detrimental impact than weekend use, while high absenteeism (absences) is directly correlated with lower performance. Conversely, support systems like extra paid classes and internet access show a positive influence. Synthesizing these findings allows for the creation of a powerful early warning system: students with a G2 below 10, multiple past failures, or high absences can be automatically flagged for immediate academic support, while those with good grades but high behavioral risks (like weekday drinking) can be targeted for counseling. Furthermore, the model can identify students on improving or declining trajectories between G1 and G2, enabling educators to provide either positive reinforcement or preemptive support before final failure occurs.

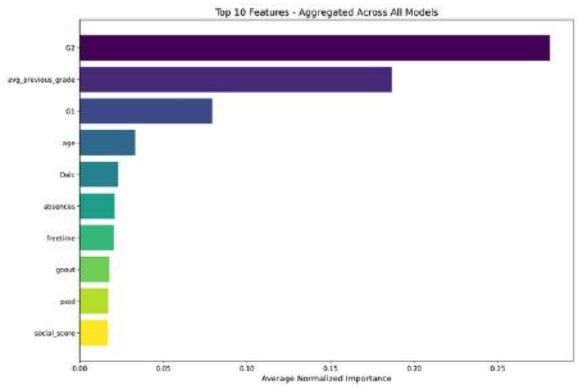

Fig.4 Feature importance aggregated

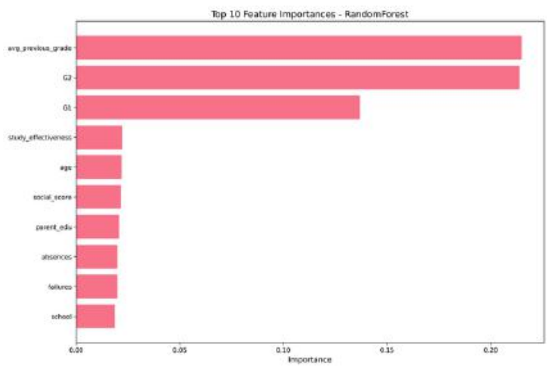

Fig.5 Feature importance Random Forest

VI. LIMITATIONS AND FUTURE WORK

It is crucial to acknowledge the limitations of this study. The model is trained on historical data (2005-2006) from a specific Portuguese context, which may not generalize perfectly to other cultures or modern school systems. The dataset also has an urban bias (77% of students). Furthermore, the model can only capture the features provided. It cannot predict sudden external life events (e.g., family crises, health issues) or internal changes in student motivation that are not reflected in their past behavior.

Finally, while our ensemble model is robust, the Random Forest model performed better on its own. Future work should explore more advanced ensemble techniques, such as stacking, or perform a more exhaustive hyperparameter search (e.g., Bayesian optimization) to further enhance performance. Nonetheless, the 76.92% accuracy of the Random Forest model provides a powerful and highly interpretable tool for immediate use by educators to identify and support at-risk students.

VII. CONCLUSION

This study highlights the critical role of data-driven approaches in improving academic outcomes in higher education. By benchmarking multiple machine learning models for student performance prediction, institutions can better understand which techniques provide the most accurate and reliable insights. The integration of a weighted fusion approach further demonstrates how combining model strengths can enhance prediction quality and support more informed decision-making. These findings contribute to the growing field of educational data mining by offering practical guidance for universities seeking to identify at-risk students early, personalize learning pathways, and improve overall academic planning. As higher education continues to adopt digital and analytics-based solutions, the use of effective predictive models will remain essential for promoting student success and supporting evidence-based teaching and learning practices.

REFERENCES

[1] Amjad, S., et al., Data mining techniques to analyze the impact of social media on academic performance of high school students. Wireless Communications and Mobile Computing, 2022. 2022(1): p. 9299115.

[2] Sarker, I.H., Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2021. 2(3): p. 160.

[3] Hilbert, S., et al., Machine learning for the educational sciences. Review of Education, 2021. 9(3): p. e3310.

[4] Tang, Y., et al., Introduction to focus issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2020. 30(6).

[5] Osamor, V.C. and A.F. Okezie, Enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis. Scientific Reports, 2021. 11(1): p. 14806.

[6] Mohammed, A. and R. Kora, A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences, 2023. 35(2): p. 757-774.

[7] Mazhar, T., et al., Predictive analytics in education-enhancing student achievement through machine learning. Social Sciences & Humanities Open, 2025. 12: p. 101824.

[8] Villegas, W., et al. Machine learning models for academic performance prediction: interpretability and application in educational decision-making. in Frontiers in Education. 2025. Frontiers.

[9] Asselman, A., M. Khaldi, and S. Aammou, Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 2023. 31(6): p. 3360-3379.

[10] Kumari, P., P.K. Jain, and R. Pamula. An efficient use of ensemble methods to predict students academic performance. in 2018 4th International Conference on Recent Advances in Information Technology (RAIT). 2018. IEEE.

[11] Singh, R. and S. Pal, Machine learning algorithms and ensemble technique to improve prediction of students performance. International Journal of Advanced Trends in Computer Science and Engineering, 2020. 9(3).

[12] Sekeroglu, B., K. Dimililer, and K. Tuncal. Student performance prediction and classification using machine learning algorithms. in Proceedings of the 2019 8th international conference on educational and information technology. 2019.

[13] Ahmed, E., Student performance prediction using machine learning algorithms. Applied computational intelligence and soft computing, 2024. 2024(1): p. 4067721.

[14] Chen, Y. and L. Zhai, A comparative study on student performance prediction using machine learning. Education and Information Technologies, 2023. 28(9): p. 12039-12057.