Article

A Multi-Algorithm Fusion Approach to Predicting Student Performance

Dongmei Bao[1] and Lkhagva Ariunjargal[2,*]

[1] College of Artificial Intelligence and Big Data, Hulunbuir University, Hulunbuir, China
[2] Department of Educational Studies and Methodology, School of Educational Studies, Mongolian National University of Education, Ulaanbaatar, Mongolia
* Correspondence: Lkhagva Ariunjargal, Department of Educational Studies and Methodology, School of Educational Studies, Mongolian National University of Education, Ulaanbaatar, Mongolia

Abstract: In contemporary blended learning environments that seamlessly integrate online and offline educational components, the ability to accurately predict students' academic performance using formative learning data has emerged as a critical factor in improving overall teaching quality. However, existing approaches often face significant challenges, particularly regarding the low accuracy and poor generalization capabilities typically associated with single prediction algorithms in student performance forecasting. To systematically address these persistent issues, this study constructs a robust multi-algorithm fusion model for performance prediction based on the Stacking ensemble learning framework. Specifically, this advanced model employs Random Forest, XGBoost, and LightGBM as foundational base learners to capture diverse data patterns, while utilizing Ridge Regression as the meta-learner to prevent overfitting. This hierarchical architecture achieves highly accurate course performance predictions through comprehensive secondary learning across multiple heterogeneous models. To rigorously validate the proposed methodology, the study utilized comprehensive grading data from the 'C Language Programming' course—taught simultaneously through online and offline modalities to computer-related majors at Hulunbuir University—as the primary experimental dataset. The comprehensive experimental results conclusively demonstrate that the proposed multi-algorithm fusion model significantly outperforms traditional single algorithms in terms of overall regression accuracy and predictive stability. Consequently, this innovative approach provides highly reliable predictive support and actionable insights for universities, enabling administrators and educators to conduct precise, data-driven teaching evaluations and implement proactive academic quality monitoring systems effectively.

Keywords: performance prediction; ensemble learning; data mining; blended learning; machine learning

Received: 29 April 2026
Revised: 16 June 2026
Accepted: 26 June 2026
Published: 30 June 2026

1. Introduction

In the era of 'Internet Plus', the integration of artificial intelligence with education has significantly transformed teaching and learning methodologies [1, 2]. The adoption of blended learning models, which combine online and offline teaching, has become increasingly prevalent in Chinese universities. These institutions are actively working to merge information technology with traditional course delivery methods to enhance educational outcomes. However, challenges persist, such as the physical separation between teachers and students in online environments, which often leads to reduced interaction and engagement. This lack of interaction has been identified as a key factor contributing to low student participation and a widening gap in academic performance among learners. Addressing these issues requires innovative approaches to identify students facing academic difficulties with precision, leveraging the vast amounts of data

generated during their learning processes. By doing so, educators can implement targeted strategies to improve teaching quality and foster better learning outcomes.

Academic performance prediction has emerged as a critical tool for identifying students at risk of underperforming and for enabling tailored support mechanisms. Various predictive models have been developed using algorithms such as decision trees, support vector machines, and neural networks. For example, some researchers have explored advanced neural network architectures to forecast student grades, while others have experimented with combining multiple models to enhance predictive accuracy. Despite these advancements, single-algorithm approaches often face limitations due to their dependency on specific dataset characteristics. This dependency can result in inconsistent performance, particularly when applied to diverse educational contexts [3]. To overcome these challenges, multi-algorithm fusion strategies have gained traction. These strategies combine the strengths of different algorithms to achieve higher accuracy and robustness. For instance, within a Stacking framework, integrating algorithms like K-nearest neighbours, random forests, neural networks, and decision trees as base learners, and using a meta-learner to optimise their combination, has shown superior predictive performance. Such approaches have demonstrated their effectiveness not only in education but also in fields like medicine and engineering, where they leverage the complementary strengths of heterogeneous models to achieve more reliable outcomes.

While the effectiveness of Stacking fusion strategies has been validated in prior research, there remains a need for further exploration into the optimal selection of base learners and meta-learners, particularly for process-oriented data in hybrid learning environments. This study aims to address this gap by constructing a multi-algorithm fusion model specifically designed for predicting academic performance in such scenarios. By integrating multiple heterogeneous base learners and employing a meta-learner to refine the fusion strategy, the model seeks to enhance both predictive accuracy and generalisation capabilities [4]. This approach is expected to provide more reliable support for data-driven teaching evaluation and academic quality monitoring in higher education institutions. Ultimately, the goal is to enable educators to make informed decisions based on robust predictive insights, thereby improving the overall effectiveness of teaching and learning processes in universities.

2. Relevant Theories and Techniques

2.1. Data-Driven Models

This study selected seven commonly used regression models as baselines, encompassing a diverse range of methodologies, including linear models, tree-based models, ensemble learning techniques, and kernel-based approaches. Linear regression and ridge regression are employed as foundational linear models, with ridge regression addressing the issue of multicollinearity through the application of L2 regularisation, thereby enhancing model stability and interpretability. Decision trees, as a non-linear model, are adept at capturing complex interactions among features; however, their inherent instability when relying on a single tree limits their generalisation capabilities. To address this, Random Forests aggregate multiple decision trees using the Bagging technique, which improves predictive accuracy and reduces overfitting. Support Vector Regression (SVR) introduces kernel functions to effectively model non-linear relationships, although its computational inefficiency on large-scale datasets poses a significant limitation. XGBoost and LightGBM, both high-performance ensemble models based on the gradient boosting framework, offer advanced capabilities. XGBoost incorporates regularisation techniques and parallel computing to enhance model robustness and scalability, while LightGBM employs histogram-based algorithms and one-sided gradient sampling, which significantly improve training speed and memory efficiency [5]. These models collectively represent a spectrum of regression paradigms, ranging from linear to highly non-linear approaches, each with distinct advantages and limitations. A detailed comparison of their core principles, strengths, and weaknesses is

summarised in Table 1, providing a comprehensive overview of their applicability in various scenarios.

Table 1. Comparison of Regression Models

Model	Core Principles	Key Advantages	Limitations
Linear Regression	Fitting a linear relationship between features and the target	Simple and highly interpretable	Unable to handle non-linear relationships
Ridge Regression	Linear regression + L2 regularisation	Mitigates multicollinearity and controls model complexity	Remains a linear model, with limited non-linear capabilities
Decision Trees	Recursively partitions the feature space, outputting the mean of the leaf nodes	Captures non-linear interactions	Single trees are prone to overfitting and lack stability
Random Forests	Bagging: ensemble of multiple decision trees with random feature selection	Resistant to overfitting with good generalisation performance	Models are large and training is slow
SVR	Finding a regression hyperplane in a high-dimensional feature space that tolerates an error ε	Handles non-linearity via kernel functions	Training efficiency is low with large-scale data
XGBoost	Gradient Boosting framework: iteratively fits residuals, including regularisation	High accuracy, supports parallel processing, and handles missing values	Sensitive to parameter tuning
LightGBM	Histogram-based Gradient Boosting: one-sided gradient sampling	Fast training and low memory usage	May overfit on small datasets

Stacking, also known as stacked generalisation, represents a sophisticated hierarchical ensemble learning framework designed to enhance predictive performance by combining the strengths of multiple base learners [4]. The fundamental concept involves using the predictions generated by diverse base models as new input features, which are subsequently processed by a meta-learner to produce the final predictions. This layered approach allows Stacking to mitigate the 'weakest link effect' often observed in single-model frameworks, where the limitations of an individual model can disproportionately affect overall performance. By leveraging the complementary strengths of heterogeneous base learners, Stacking achieves improved accuracy and robustness in predictions. For instance, linear models may excel in capturing global trends, while tree-based models are better suited for identifying local interactions; combining these through Stacking enables a more holistic modelling approach. Furthermore, the meta-learner, typically a high-capacity model, is trained to optimise the integration of base learner outputs, ensuring that the ensemble capitalises on the unique contributions of each component. The structural design of the Stacking model, as illustrated in Figure 1, underscores its ability to synthesise diverse predictive insights into a unified framework, making it a powerful tool for tackling complex regression tasks.

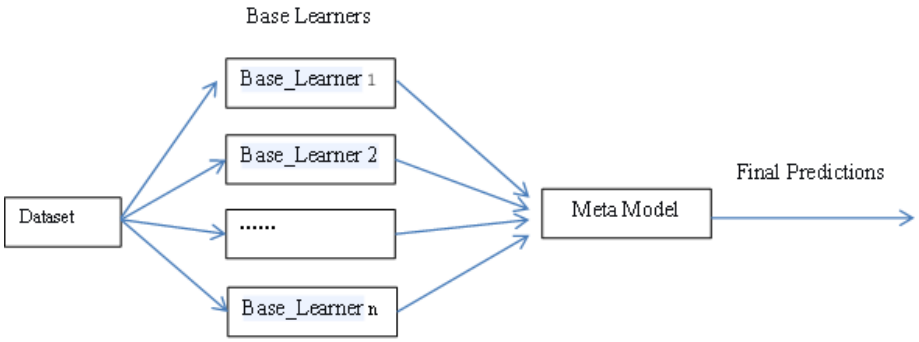

Figure 1. Structure of the Stacking Model

2.2. Model Evaluation Metrics

To objectively evaluate the predictive performance of each model, this study employs three widely recognized regression evaluation metrics: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R2). These metrics are essential for assessing the accuracy and reliability of predictive models in various fields, including machine learning, statistics, and data analysis. Each metric provides unique insights into the model's performance, enabling a comprehensive evaluation. MSE quantifies the average squared difference between predicted and actual values, emphasizing larger errors due to its squaring effect. MAPE, on the other hand, offers a percentage-based measure of error, making it particularly useful for comparing models across datasets with varying scales. Lastly, R2 evaluates the proportion of variance in the dependent variable that is explained by the model, serving as a measure of its explanatory power. Together, these metrics form a robust framework for model evaluation, ensuring that both accuracy and goodness of fit are adequately addressed [6].

Mean Squared Error (MSE) is a fundamental metric used to measure the average squared differences between predicted and actual values. It is particularly sensitive to large errors due to the squaring operation, which amplifies the impact of outliers. This characteristic makes MSE an effective tool for identifying models that consistently produce predictions close to the actual values. A smaller MSE value indicates higher predictive accuracy, as it reflects minimal deviation between the predicted and observed data points. The formula for MSE is critical in regression analysis and is often used in conjunction with other metrics to provide a balanced evaluation of model performance. By focusing on the squared differences, MSE ensures that both the magnitude and direction of errors are accounted for, making it a comprehensive measure of predictive accuracy.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Mean Absolute Percentage Error (MAPE) is a widely used metric that provides an intuitive understanding of a model's predictive accuracy by expressing errors as a percentage of the actual values. This percentage-based approach makes MAPE particularly useful for comparing models across datasets with different scales or units. A lower MAPE value indicates better model performance, as it signifies smaller relative errors. However, it is important to note that MAPE can be sensitive to very small actual values, which may disproportionately inflate the error percentage. Despite this limitation, MAPE remains a popular choice for evaluating forecasting models due to its simplicity and interpretability. By focusing on the relative magnitude of errors, MAPE allows researchers to assess the practical implications of model predictions in real-world scenarios [7].

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

The Coefficient of Determination (R2) is a key metric used to evaluate the goodness of fit of a regression model. It measures the proportion of variance in the dependent variable that is explained by the independent variables in the model [8]. R2 values range

from negative infinity to 1, with values closer to 1 indicating a better fit. A high R2 value suggests that the model effectively captures the underlying patterns in the data, while a low R2 value may indicate that the model fails to account for significant variability. It is important to interpret R2 in the context of the specific dataset and research objectives, as a high R2 does not necessarily imply causation or predictive accuracy. By providing a measure of explanatory power, R2 helps researchers assess the overall effectiveness of their models in representing the observed data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

3. Data and Methods

3.1. Data Sources and Preprocessing

The data utilized in this study were derived from widely used educational platforms, including Chaoxing, PTA, and Xiji, which support the 'C Language Programming' course at Hulunbuir University. A comprehensive dataset was compiled, encompassing the complete learning records of 349 students enrolled in the course. The raw data fields included a variety of academic performance indicators, such as online engagement metrics from the Chaoxing platform, scores from offline assignments, mid-term examination results, practical assessment marks, and final examination scores [5]. After a rigorous feature selection process, 25 predictive features were identified as being most relevant for the study's objectives. These features were carefully chosen to ensure they provided meaningful insights into the students' academic performance and engagement patterns, thereby enhancing the robustness of the predictive model.

To facilitate the development and evaluation of the predictive model, the dataset was divided into training and testing subsets in an 8:2 ratio. This split was performed using a fixed random seed to ensure that the results could be reproduced in subsequent analyses. The pre-processing methodology was designed to maximize the quality and reliability of the data, thereby minimizing potential biases or inaccuracies [9]. This approach ensures that the model's performance metrics accurately reflect its predictive capabilities and are not unduly influenced by data inconsistencies or noise.

(1) Data cleaning: The initial step involved a thorough examination of the dataset to identify and remove duplicate records, which could otherwise distort the analysis. Outliers were systematically detected and corrected to maintain the integrity of the data. Additionally, personal identifiers, such as student ID numbers, were anonymized to protect the privacy of the participants and comply with ethical research standards [10]. This step was critical in ensuring that the dataset was both accurate and suitable for subsequent analysis.

(2) Missing value handling: The dataset revealed instances where students had not submitted assignments or laboratory work, resulting in corresponding marks being recorded as 0. Rather than imputing these missing values, they were retained as valid data points [11]. This decision was based on the premise that such omissions are indicative of insufficient academic engagement, which is a key factor in the context of academic warning systems. By preserving these values, the dataset more accurately reflects the students' overall academic behavior and engagement levels.

(3) Feature and label definition: The study utilized the 25 formative assessment fields as predictor features, which were carefully selected to capture a comprehensive range of academic activities and performance metrics. The final examination marks were designated as the target label, serving as the primary outcome variable for the predictive model. This approach ensures that the model is trained to identify patterns and relationships between formative assessments and final academic performance, thereby providing actionable insights for educational interventions.

3.2. Design of a Multi-Algorithm Ensemble Model

This study employs the Stacking ensemble learning framework to construct a robust model for predicting student grades. Stacking is a sophisticated ensemble method that

integrates multiple heterogeneous base learners within a hierarchical two-layer structure. The first layer consists of diverse base models that independently learn from the data, while the second layer employs a meta-learner to determine the optimal way to combine the outputs of these base models. This approach effectively addresses the 'weakest link effect,' a common limitation of single-model approaches, by leveraging the strengths of multiple algorithms. The overall process is visually represented in Figure 2, which outlines the multi-algorithm fusion method. By combining the predictive capabilities of different algorithms, Stacking enhances the model's generalisation ability and reduces the risk of overfitting, making it a powerful tool for complex prediction tasks [12].

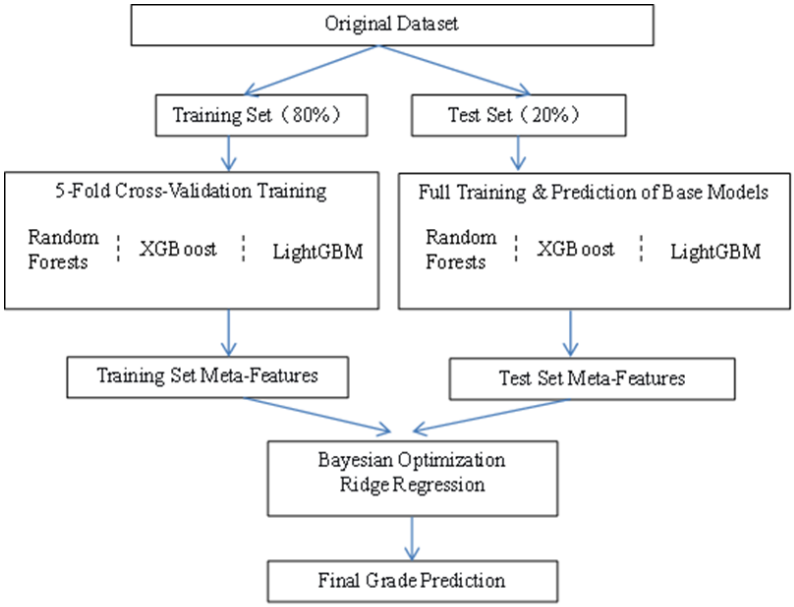

Figure 2. Multi-algorithm Fusion Method

First layer (base learners): In this study, Random Forest, XGBoost, and LightGBM are selected as the base learners due to their proven effectiveness in handling structured data. Each base learner is configured with uniform parameters to ensure a fair comparison of their individual performance and their contribution to the ensemble model. Specifically, Random Forest is set with n_estimators=100, while XGBoost and LightGBM are both configured with n_estimators=100 and learning_rate=0.1. To mitigate the risk of overfitting and enhance the reliability of the predictions, 5-fold cross-validation is employed. This involves randomly dividing the training dataset into five equal parts. During each iteration, four parts are used to train the base models, while the remaining part is used for prediction. This process is repeated five times, ensuring that each subset serves as a validation set once. The predicted values generated by each base model across all iterations are then aggregated to form a new set of meta-features. These meta-features serve as inputs for the second layer of the ensemble framework, enabling the meta-learner to capture complex patterns and interactions that may not be evident in the original feature space.

Second layer (meta-learner): Ridge regression is employed as the meta-learner in this study due to its ability to control model complexity through L2 regularisation. This regularisation technique effectively addresses multicollinearity issues, ensuring that the model remains stable and interpretable. To optimise the performance of the ridge regression model, the Optuna framework is utilised for Bayesian optimisation of the regularisation parameter α. The search range for α is set between 1e-3 and 10, and the optimisation objective is defined as the average R2 value obtained through 5-fold cross-validation. A total of 30 trials are conducted to identify the optimal value of α. Once the best parameter is determined, the ridge regression model is trained on the complete set of

meta-features, ensuring that it fully leverages the information provided by the base learners [13]. This two-layer structure allows the ensemble model to achieve a balance between bias and variance, resulting in improved predictive performance.

During the testing phase, the ensemble model follows a systematic procedure to generate predictions. First, the base models in the first layer are retrained using the entire training dataset to ensure that they capture all available information. Next, these retrained base models are used to generate meta-features for the test set samples. These meta-features, which encapsulate the predictions of the base models, are then fed into the trained ridge regression meta-model. The meta-model processes these inputs to produce the final predicted scores. To evaluate the effectiveness of the ensemble model, its performance is compared against that of the default XGBoost model, which demonstrated the highest accuracy among the individual base learners. This comparison highlights the advantages of the ensemble approach in terms of predictive accuracy and robustness, showcasing its potential for applications in educational data analysis and beyond.

4. Experimental Results and Analysis

4.1. Comparison of Single Algorithm Performance

To validate the effectiveness of the multi-algorithm fusion model, an initial evaluation was conducted to compare the predictive performance of six individual regression models: linear regression, decision tree, random forest, SVR, XGBoost, and LightGBM. Each model was trained and tested using an identical dataset split ratio of 8:2, ensuring consistency in the evaluation process. Default parameters were applied during training and prediction to maintain uniformity across models. The evaluation metrics employed included mean squared error (MSE), coefficient of determination (R2), and mean absolute percentage error (MAPE), which collectively provide a comprehensive assessment of model accuracy and reliability [5]. The results of these evaluations are presented in Table 2, while the corresponding predicted fitting curves for each model are visually depicted in Figure 3. These visual and quantitative analyses serve as a foundation for understanding the relative strengths and weaknesses of each algorithm in predicting outcomes based on the dataset.

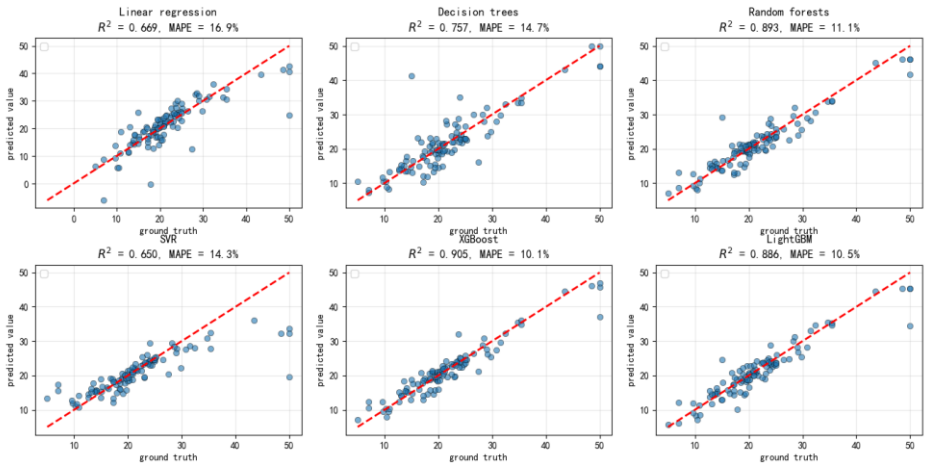

Figure 3. Model Fit Curve

Table 2. Prediction Results for Individual Models

Model	MSE	R^2	MAPE (%)
Linear regression	24.2911	0.6688	16.87
Decision trees	17.8156	0.7571	14.75
Random forests	7.8802	0.8925	11.09
SVR	25.6685	0.6500	14.30
XGBoost	6.9362	0.9054	10.15

LightGBM	8.3388	0.8863	10.47

As illustrated in Table 2, the predictive performance of the six models varied significantly. Among the tested algorithms, XGBoost demonstrated the highest accuracy, achieving an MSE of 6.9362, an R2 of 0.9054, and a MAPE of 10.15%. This was closely followed by Random Forest, which recorded an MSE of 7.8802, an R2 of 0.8925, and a MAPE of 11.09%, and LightGBM, which achieved an MSE of 8.3388, an R2 of 0.8863, and a MAPE of 10.47%. These results highlight the superior performance of tree-based ensemble models, which are particularly adept at capturing complex non-linear relationships between student grades and formative learning behaviors. In contrast, linear regression and SVR exhibited significantly lower R2 values, both below 0.67, and higher prediction errors, underscoring the limitations of linear assumptions in modeling the intricate relationships present in this dataset [12].

The decision tree model, while outperforming linear regression, exhibited notable shortcomings in comparison to the ensemble tree models. Specifically, its MSE of 17.8156 and R2 of 0.7571 were markedly inferior, reflecting its susceptibility to overfitting and limited generalization capabilities. This underscores the inherent limitations of single decision trees in handling complex datasets. Furthermore, the linear regression and SVR models struggled to capture the nuanced relationships between features and grades, as evidenced by their substantial prediction errors and low R2 values. These findings reinforce the necessity of employing more sophisticated algorithms capable of addressing non-linear dependencies and interactions within the data.

Based on the comparative analysis, the three robust tree-based models—Random Forest, XGBoost, and LightGBM—emerged as the most effective predictors for this dataset. Their superior performance across all evaluation metrics highlights their ability to model complex relationships and interactions between features and outcomes. Consequently, these algorithms were selected as the base learners for the subsequent development of the multi-algorithm fusion model, which aims to further enhance predictive accuracy by leveraging the strengths of these individual models.

4.2. Performance of Multi-Algorithm Ensemble Models

Based on the experimental results in Section 4.1, the three tree-based models with the best predictive performance—Random Forest, XGBoost, and LightGBM—were selected as base learners for the Stacking ensemble model. These models were chosen due to their demonstrated ability to handle complex data structures and their robustness in predictive tasks. Ridge Regression was employed as the meta-learner, serving as the secondary model to integrate the outputs of the base learners. The parameter settings for both the base learners and the meta-learner were consistent with those outlined in Section 3.2, ensuring methodological continuity and comparability of results. To generate meta-features for the training set, a five-fold cross-validation approach was utilized. This method not only enhances the reliability of the meta-features but also reduces the risk of overfitting by ensuring that each data point is used for both training and validation. Additionally, the regularization parameter α for Ridge Regression was optimized using the Bayesian optimization framework provided by Optuna. This optimization process allowed for the fine-tuning of the meta-learner, ensuring that it effectively captured the complementary strengths of the base learners while minimizing prediction errors.

To validate the effectiveness of the fusion model, the prediction results of the Stacking ensemble were compared with those of the default XGBoost model, which had demonstrated the best performance among the individual algorithms in prior evaluations. This comparison was conducted to assess whether the ensemble approach could achieve superior predictive accuracy and stability. The results of this comparative analysis are presented in Table 3, which highlights the performance metrics of both models. By juxtaposing the outcomes, the study aimed to provide a clear and quantitative evaluation of the benefits of employing a multi-algorithm ensemble strategy over relying on a single algorithm, even one as robust as XGBoost.

Table 3. Predictive Performance of the Ensemble Model Versus the Optimal Individual Model

Model	MSE	R^2	MAPE (%)
XGBoost	6.9362	0.9054	10.15
Stacking	6.5710	0.9104	9.94

As shown in Table 3, the multi-algorithm stacking model developed in this study achieved the best predictive performance across all evaluated metrics. Specifically, the Mean Squared Error (MSE) was reduced to 6.5710, the R-squared (R2) value increased to 0.9104, and the Mean Absolute Percentage Error (MAPE) decreased to 9.94%. When compared to XGBoost, which was the top-performing individual algorithm, the Stacking model demonstrated notable improvements. The R2 value increased from 0.9054 to 0.9104, reflecting an enhancement of 0.0046, which indicates a better fit of the model to the data. The MSE decreased from 6.9362 to 6.5710, representing a reduction in prediction error by approximately 5.27%. Similarly, the MAPE was reduced from 10.15% to 9.94%, marking a decrease of 0.21 percentage points. These improvements underscore the efficacy of the ensemble approach in leveraging the strengths of multiple algorithms to achieve superior predictive accuracy and reliability.

The experimental results provide strong evidence that a multi-algorithm fusion strategy can significantly enhance the accuracy and stability of performance prediction. By integrating three heterogeneous tree-based models—Random Forest, XGBoost, and LightGBM—through the Stacking framework, the ensemble model effectively capitalizes on the unique strengths of each base learner. Ridge Regression, employed as the meta-model, plays a critical role in secondary learning by synthesizing the outputs of the base learners into a cohesive and optimized prediction. This approach mitigates the prediction bias that individual models may exhibit on specific samples, thereby improving overall model robustness. The use of Bayesian optimization for fine-tuning the meta-learner further ensures that the ensemble model achieves optimal performance. These findings highlight the potential of ensemble learning techniques to address the limitations of single-model approaches, particularly in complex predictive tasks where data variability and noise can pose significant challenges [5].

5. Conclusions

This study addressed the problem of unstable predictive performance commonly observed in single-algorithm approaches to student grade prediction in blended learning environments by developing a multi-algorithm fusion prediction model based on the Stacking ensemble framework. The proposed model combined Random Forest, XGBoost, and LightGBM as base learners and employed Bayesian-optimised Ridge Regression as the meta-learner, thereby constructing a two-layer ensemble architecture for continuous prediction of student course scores. Through the complementary integration of multiple learning algorithms, the model was designed to capture heterogeneous patterns in student learning behaviour and assessment-related data more effectively than any individual method operating in isolation.

Experimental evaluation using real learning data from 349 students enrolled in the "C Language Programming" course at Hulunbuir University showed that the proposed Stacking fusion model achieved superior overall performance relative to the best-performing single model, namely XGBoost. Across the three regression evaluation metrics, including MSE, R2, and MAPE, the fusion model consistently demonstrated improved predictive capability. In particular, the R2 value reached 0.9104, indicating strong explanatory power with respect to the variance in final course grades. At the same time, MSE was reduced by approximately 5.27%, and MAPE decreased to 9.94%, further confirming that the ensemble strategy enhanced both prediction precision and error control. These results indicate that the fusion of diverse learners can effectively improve model stability, reduce the limitations associated with reliance on a single algorithm, and strengthen the practical reliability of grade prediction.

From an application perspective, the proposed method offers meaningful support for data-driven teaching evaluation, academic quality monitoring, and early identification of student performance trends in higher education contexts. Its relatively strong generalisation ability and predictive accuracy suggest that it may serve as a useful analytical tool for assisting instructors and administrators in understanding learning outcomes and optimising instructional interventions.

Nevertheless, several limitations should be acknowledged. In the present study, the base learners adopted fixed parameter settings and did not undergo systematic hyperparameter tuning, which may have constrained the full performance potential of the ensemble. In addition, the dataset was derived from a single course at a single institution, and therefore the external validity of the model across different disciplines, course structures, and educational settings remains to be further examined. Future research should expand the data sources to include multiple courses and institutions, investigate automated hyperparameter optimisation strategies for the base learners, and explore the integration of deep learning methods with ensemble learning frameworks in order to further improve the accuracy, robustness, and adaptability of student grade prediction models.

Funding: 2024 Research Project of Hulunbuir College "Research on Grade Prediction Model Based on Multi-Algorithm Fusion" (Project No.: 2024XJCG21); 2024 Annual Project of the "14th Five-Year" Plan for Educational Science Research of Inner Mongolia Autonomous Region (Project No.: NGJGH2024628).

References

1. L. Zhang, N. F. Ibrahim, and Z. Zainol, "LSTM-Based Student Performance Prediction Model Incorporating Peer Effects and Explained Through SHAP," *Journal of Engineering*, vol. 2025, no. 1, p. 5731919, 2025.
2. V. Balachandar and K. Venkatesh, "A multi-dimensional student performance prediction model (MSPP): An advanced framework for accurate academic classification and analysis," *MethodsX*, vol. 14, p. 103148, 2025.
3. A. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting student performance in higher education institutions using decision tree analysis," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, pp. 26–31, 2018.
4. A. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques," *Journal of Computing in Higher Education*, pp. 1–39, 2025.
5. M. Kumar, K. Bajaj, B. Sharma, and S. Narang, "A comparative performance assessment of optimized multilevel ensemble learning model with existing classifier models," *Big Data*, vol. 10, no. 5, pp. 371–387, 2022.
6. N. Sateesh, P. Srinivasa Rao, and D. Rajya Lakshmi, "Optimized ensemble learning-based student's performance prediction with weighted rough set theory enabled feature mining," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 7, p. e7601, 2023.
7. X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 241–258, 2020.
8. Y. Wu, "From ensemble learning to deep ensemble learning: A case study on multi-indicator prediction of pavement performance," *Applied Soft Computing*, vol. 166, p. 112188, 2024.
9. I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access*, vol. 10, pp. 99129–99149, 2022.
10. B. Tang, S. Li, and C. Zhao, "Predicting the performance of students using deep ensemble learning," *Journal of Intelligence*, vol. 12, no. 12, p. 124, 2024.
11. O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.
12. G. Singh, J. Gómez-Luna, G. Mariani, G. F. Oliveira, S. Corda, S. Stuijk, et al., "Napel: Near-memory computing application performance prediction via ensemble learning," in *Proceedings of the 56th Annual Design Automation Conference 2019*, pp. 1–6, June 2019.
13. R. Guo, D. Fu, and G. Sollazzo, "An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree," *International Journal of Pavement Engineering*, vol. 23, no. 10, pp. 3633–3646, 2022.