Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

A Comparative Analysis of Machine Learning Models for Predicting Student Performance: Evaluating the Impact of Stacking and Traditional Methods

Gregorius Airlangga[1]
[1]Atma Jaya Catholic University of Indonesia, Indonesia
[1]gregorius.airlangga@atmajaya.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 16-09-2024
Accepted: 23-09-2024
Published: 04-10-2024

Keywords:
Student performance prediction; Machine learning models; Ensemble learning; Stacking regressor; Educational data mining.

ABSTRACT

This study investigates the application of machine learning models to predict student performance using socio-economic, demographic, and academic factors. Various models were developed and evaluated, including Linear Regression, Random Forest, Gradient Boosting, XGBoost, LightGBM, Support Vector Regressor, and a Stacking Regressor. The models were assessed using key evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R^2), Mean Squared Log Error (MSLE), and Mean Absolute Percentage Error (MAPE). The Support Vector Regressor demonstrated the best overall performance, with an MAE of 4.3091, RMSE of 5.4110, and an R^2 of 0.8685, surpassing even the more complex ensemble models. Similarly, Linear Regression achieved strong results, with an MAE of 4.3154 and R^2 of 0.8685. In contrast, the Stacking Regressor, while effective, did not significantly outperform its base models, achieving an MAE of 4.5340 and R^2 of 0.8563, highlighting that greater model complexity does not necessarily lead to better predictive power. The analysis also revealed that MAPE was highly sensitive to outliers in the dataset, indicating the need for robust data preprocessing to handle extreme values. These results suggest that, in educational data mining, simpler models can often match or exceed the performance of more complex methods. Future research should investigate advanced ensembling strategies and feature engineering techniques to further enhance the accuracy and reliability of student performance predictions.

INTRODUCTION

Understanding the factors that influence student performance is crucial for educational institutions, policymakers, and educators to devise strategies that can enhance learning outcomes (D'iez et al., 2020). Academic performance is not solely determined by innate abilities; it is influenced by a myriad of factors, including socio-economic background, parental education, and learning environment. In recent years, there has been growing interest in using data-driven approaches to explore these influences (Davids, 2020; Marks & O'Connell, 2021; Razzaq et al., 2024). With the increasing availability of educational data, machine learning has emerged as a powerful tool to predict student performance, offering new insights into how various factors interplay to affect learning outcomes (Hasan et al., 2020; Hooda et al., 2022; Kabathova & Drlik, 2021). The relationship between student performance and various socio-economic and demographic factors has been extensively studied in the past. For example, (Ludeke et al., 2021) found a significant correlation between parental education level and student academic success, indicating that students with highly educated parents tend to perform better in standardized tests. Similarly, (Rodr'iguez-Hernández et al., 2020) analyzed the impact of socio-economic status and concluded that students from higher-income families generally have access to better educational resources, leading to improved academic outcomes. Another aspect frequently explored is the influence of school-related factors, such as lunch programs and test preparation courses. (Roberts et al., 2022) suggested that adequate nutrition through standard lunch programs positively affects cognitive function, while (Hooda et al., 2022) emphasized the role of structured test preparation courses in enhancing student test scores.

While traditional statistical methods like regression analysis have been used in earlier studies to identify correlations and predictive factors, there is a growing shift towards leveraging more sophisticated machine learning models to enhance the prediction accuracy of student performance. Recent studies by (Badal & Sungkur, 2023; Hashim et al., 2020; Ofori et al., 2020) have demonstrated the efficacy of machine learning algorithms such as Random Forest, Support Vector Machines, and Neural Networks in predicting student outcomes with higher precision than traditional methods. However, these studies often focus on a limited set of predictive features or use relatively simple models. There is still a research gap in exploring how complex ensemble and stacking models can be applied to student performance data, incorporating a wider range of factors for more nuanced and accurate predictions. The urgency of this research stems from the pressing need to improve educational outcomes in diverse learning environments. Educational institutions are increasingly under pressure to identify at-risk students early and implement targeted interventions. However, traditional methods of assessing student performance are often reactive and fail to account for the intricate

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

interplay of various factors that influence learning. By employing advanced machine learning techniques, this research aims to develop predictive models that can proactively identify students who may be at risk of underperforming, allowing for timely and more personalized interventions. Furthermore, understanding the specific factors that significantly influence student performance can inform policy decisions, curriculum design, and resource allocation, ultimately contributing to more equitable and effective education systems.

Current advancements in predictive analytics have seen the integration of machine learning models, such as Random Forest, Gradient Boosting, and Support Vector Regression, in educational data mining (Domladovac, 2021; Syed Mustapha, 2023; Yaugci, 2022). These models have shown promise in handling complex, non-linear relationships between predictive features and student performance outcomes. Ensemble learning methods have gained traction due to their ability to combine the strengths of multiple base models to improve predictive accuracy. However, despite these advancements, challenges persist in selecting the optimal combination of models and preprocessing techniques to maximize predictive performance (Alawsi et al., 2022). Moreover, while prior research has incorporated machine learning models to some extent, the use of more advanced ensemble techniques like Stacking Regressors remains relatively unexplored in the context of educational data (Bento et al., 2021). The existing body of research has provided valuable insights into the factors influencing student performance and the application of machine learning models in educational settings. However, there are notable gaps that this study aims to address. Firstly, many studies have utilized relatively simple models or focused on a narrow set of predictors, potentially overlooking the complex interactions between various socio-economic, demographic, and academic factors. Secondly, there is a lack of comprehensive evaluations of advanced ensemble models, such as stacking and blending techniques, which have the potential to outperform single-model approaches by combining multiple predictive models to capture different aspects of the data. Lastly, while some research has explored the impact of specific factors like parental education and test preparation, there is a need for a more holistic approach that considers a broader range of variables in a single predictive framework.

The primary goal of this research is to develop and evaluate advanced machine learning models that can accurately predict student performance based on a diverse set of features, including gender, ethnicity, parental education level, lunch program participation, and test preparation status. By applying a range of regression models, including Random Forest, Gradient Boosting, XGBoost, LightGBM, Support Vector Regression, and Stacking Regressors, this study aims to identify the most effective predictive framework for understanding how different factors contribute to student success. Additionally, the research seeks to provide actionable insights into which factors have the most significant impact on student performance, offering a data-driven foundation for educational policy and intervention strategies. This research contributes to the field of educational data mining by applying a comprehensive set of machine learning models to student performance prediction. The novel contribution lies in the use of advanced ensemble techniques, particularly the Stacking Regressor, which combines multiple base models to enhance predictive accuracy. By systematically evaluating these models using a range of performance metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R2), Mean Squared Log Error (MSLE), and Mean Absolute Percentage Error (MAPE), the study provides a nuanced understanding of the predictive power of different algorithms. Furthermore, this research extends the current understanding of how various socio-economic and demographic factors interact to influence student outcomes, offering a more holistic predictive framework compared to previous studies. The remainder of this article is structured as follows: Section 2 details the literature survey, Section 3 describes dataset, and the preprocessing steps applied to prepare the data for modeling, including feature engineering, handling of categorical variables, and scaling of numerical features. In addition, we also outline the machine learning models used in this study, including the hyperparameter tuning process and the stacking technique employed to enhance predictive performance. Section 3 presents the experimental results, comparing the performance of different models using the selected evaluation metrics. This section also includes an in-depth discussion of the findings, highlighting the most influential factors affecting student performance. Finally, Section 4 concludes the study, summarizing the key contributions, implications for educational practice, and potential avenues for future research.

LITERATURE REVIEW

Predicting student performance is crucial for identifying at-risk students and implementing targeted interventions. This prediction can help educators and policymakers enhance educational strategies and outcomes. Although various factors influence student performance, the integration of advanced machine learning techniques offers new opportunities to accurately capture and analyze these factors' complex relationships. This literature review examines key studies on socio-economic, demographic, and school-related factors affecting performance, explores current machine learning applications, and identifies gaps that this research aims to address. Numerous studies have explored the relationship between socio-economic status, demographic factors, and academic outcomes. Parental education is frequently cited as a significant determinant. For example, (Tan et al., 2020) demonstrated that students with parents holding higher educational qualifications generally achieve better academic results due to the availability of resources and a supportive home environment conducive to learning. This finding suggests that students' academic performance can be partially predicted by considering parental educational background. Similarly, (Hultberg et al., 2021) found that students from higher-income families benefit from enhanced learning resources, extracurricular

Brilliance Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 4, Number 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

activities, and private tutoring, all of which contribute to improved academic performance.

Ethnicity is another demographic factor that has been shown to affect student outcomes. (Blanden, 2020) observed disparities in performance among different ethnic groups, often tied to socio-economic inequalities and differential access to quality education. These disparities underscore the need for predictive models that consider ethnicity and socio-economic status to understand the multifaceted nature of academic success fully. This research seeks to integrate these variables into a comprehensive model, addressing how they jointly influence performance in educational settings. Parental involvement is also a key factor in academic success. (Goshin et al., 2021) demonstrated that students with active parental support, including monitoring of homework and engagement in school activities, tend to achieve higher test scores. Furthermore, parents with higher educational backgrounds are often more equipped to assist with academic challenges, reinforcing the connection between parental education and student performance. The present research incorporates these findings by including parental education as a predictive feature, aiming to quantify its direct and indirect effects on student test scores.

Test preparation has been recognized as a crucial determinant of student performance. (Strelan et al., 2020) highlighted the positive impact of structured test preparation courses on students' test scores. Their study found that students who completed test preparation programs, especially those from lower socio-economic backgrounds, demonstrated substantial improvements in academic performance. By including test preparation status as a feature, our research aims to capture its role in predicting students' academic outcomes, offering insights into the effectiveness of such programs. Nutrition is a foundational factor in cognitive development and academic success. (Cohen et al., 2021) found that participation in school lunch programs, particularly those offering nutritious meals, is associated with better student performance. This finding is critical for predictive modeling as it suggests that nutritional status, often linked with lunch program participation, can be a significant predictor of academic outcomes. By incorporating lunch program participation as a variable, this research seeks to explore the extent to which nutritional status, as indicated by lunch program type, contributes to predicting student test scores.

Traditional statistical methods, while useful, often fall short in capturing the non-linear relationships and interactions between diverse factors influencing student performance. Machine learning (ML) models, particularly ensemble methods, have demonstrated superior predictive capabilities in this domain. For example, (Syed Mustapha, 2023) used Random Forest and Support Vector Regression (SVR) to predict student grades based on socio-demographic factors, finding that Random Forest outperformed SVR due to its ability to handle complex interactions. This underscores the potential of tree-based models for capturing non-linear patterns in educational data. Building on these findings, (Villar & de Andrade, 2024) employed Gradient Boosting Machines (GBM) to predict student dropout rates, achieving high predictive accuracy. These studies emphasize the effectiveness of machine learning models, particularly ensemble methods, in educational settings. However, they often focus on individual algorithms rather than leveraging the combined power of multiple models. This research aims to fill this gap by applying advanced ensemble techniques, such as stacking, to combine multiple models and improve predictive accuracy.

While single machine learning models like Random Forest and Gradient Boosting have shown effectiveness in educational data mining, the use of advanced ensemble techniques such as stacking remains underexplored. Stacking involves training several base models and using their predictions as inputs to a final model, thereby capturing diverse patterns in the data. (Xue & Niu, 2023; Yu & Liu, 2022) demonstrated that stacking models, using Random Forest, Gradient Boosting, and LightGBM as base models, outperformed individual models in predicting student grades. This suggests that stacking can improve model performance by leveraging the strengths of different algorithms, a promising direction that this research intends to explore further. Our research builds upon these findings by not only incorporating stacking techniques but also comparing their effectiveness against individual models like Random Forest, Gradient Boosting, XGBoost, and Support Vector Regression. By doing so, we aim to identify the optimal predictive framework for understanding how various socio-economic, demographic, and school-related factors jointly contribute to student performance.

METHOD

The proposed methodology is structured to predict student performance by leveraging advanced ensemble learning techniques, focusing on stacking regressors to achieve high predictive accuracy. This approach addresses the complex interactions among various socio-economic, demographic, and academic factors. The methodology is implemented using Python, with primary reliance on libraries such as scikit-learn, XGBoost, LightGBM, and pandas. The entire pipeline, encompassing data preprocessing, model training, and evaluation, is automated to ensure reproducibility and facilitate scalability for future research endeavors. This integrated approach provides a novel and mathematically rigorous solution to the problem of predicting student performance.The methodology consists of key phases: dataset preparation, data preprocessing, model development, and model evaluation, as illustrated conceptually in Figure 1.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

Dataset Preparation

The dataset utilized in this study includes various features such as gender $((X_{\text{gender}}))$, ethnicity $((X_{\text{ethnicity}}))$, parental education $((X_{\text{parental\_education}}))$, lunch status $((X_{\text{lunch}}))$, test preparation $((X_{\text{test\_prep}}))$, and subject scores, $(S_{\{math\}})$, $(S_{\{reading\}})$, $(S_{\{writing\}})$. Let (X) denote the feature matrix and $(y = S_{(math)}$ the target variable. The dataset was further enriched with additional features such as the total score $(S_{\{total\})}$ and average score $(S_{\{average\})}$ to provide a more holistic view of academic performance. The total score is defined as presented in equation 1.

$$S_{\{total\}} = S_{\{math\}} + S_{\{reading\}} + S_{\{writing\}} \tag{1}$$

and the average score is given by in the equation 2.

$$S_{\{average\}} = \frac{S_{total}}{3} \tag{2}$$

These features were engineered to capture the overall academic aptitude of the students, providing additional input to the model to understand the relationships among different subject performances. However, to prevent multicollinearity, only relevant features were included in the final feature matrix (X).

Data Preprocessing

Data preprocessing plays a crucial role in transforming raw data into a format that machine learning models can efficiently process. This stage involves several steps: encoding categorical variables, scaling numerical features, polynomial feature expansion, and optional dimensionality reduction. Categorical variables $((C))$ such as gender, ethnicity, parental education, lunch, and test preparation were converted into numerical representations using One-Hot Encoding. Let $(C = \{X_{gender}, X_{\{ethnicity\}}, X_{\{parental_condition\}}, X_{\{lunch\}}, X_{test_preparation}$. For each categorical variable (X_j) in (C), One-Hot Encoding produces a binary matrix $(E_j \in R^{n \times k_j})$, where (n) is the number of samples and (k_j) is the number of unique categories in (X_j). The concatenated encoding matrix (E) is expressed as presented as $E = [E_1, E_2, \dots, E_m]$ where (m) is the number of categorical features. This transformation results in an expanded feature matrix (X_{encoded}), ensuring that the models can interpret categorical data without assuming any ordinal relationships.

Numerical features were then standardized using the StandardScaler. For each numerical feature $(x_i \in X_n)$, the standardized value (x_i') is computed as $x_i' = \frac{x_i - \mu_{x_i}}{\sigma_{x_i}}$ where (μ_{x_i}) and (σ_{x_i}) represent the mean and standard deviation of (x_i), respectively. This scaling ensures that each feature contributes equally to the model training, facilitating model convergence and improving performance, especially for algorithms sensitive to feature scales, such as Support Vector Regression (SVR) and Gradient Boosting. To capture higher-order interactions among numerical features, polynomial feature expansion was applied. Let the original numerical feature vector $(X_n \in R^m)$. Polynomial expansion of degree $(d = 2)$ generates additional interaction terms, creating an augmented feature space $(P(X_n) \in R^{m'})$, where $(m' = \binom{m+d}{d} - 1)$. The polynomial terms include combinations of the original features, allowing the model to learn non-linear dependencies. For example, for two features $(x_1) and (x_2)$, the polynomial expansion includes terms like $(x_1^2), (x_2^2)$, and $(x_1 x_2)$. The transformed feature matrix (X_{poly}) enriches the learning capability of linear models, facilitating the capture of complex patterns in the data.

Given the expanded feature space resulting from One-Hot Encoding and polynomial expansion, dimensionality reduction was optionally employed to mitigate the curse of dimensionality. Principal Component Analysis (PCA) was utilized to project the high-dimensional data into a lower-dimensional subspace while retaining most of the variance. Let (X') be the standardized feature matrix, PCA decomposes (X') into a set of orthogonal components (Z) such that $[Z = X'W]$ where (W) is a matrix whose columns are the eigenvectors of the covariance matrix of (X'). The principal components (Z) represent the directions of maximum variance, reducing the dimensionality while preserving essential information, enhancing model performance, and reducing computational complexity.

Model Development

The core contribution of this research lies in developing an advanced ensemble model, specifically a Stacking Regressor, to enhance predictive accuracy. The model development process begins with the establishment of individual base models, followed by the construction of the stacking ensemble. Baseline models include Linear Regression, Support Vector Regressor (SVR), Random Forest Regressor, Gradient Boosting Regressor, XGBoost Regressor, and LightGBM Regressor. Each of these models is trained to minimize its respective loss function. For example, the Random Forest Regressor minimizes the Mean Squared Error (MSE) across its ensemble of decision trees as presented in equation 4.

Brilliance
Research of Artificial Intelligence
E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2 \tag{4}$$

where (y_i) is the true target value, and $(\widehat{y_i})$ is the predicted value. Each decision tree in the ensemble is trained on a bootstrap sample of the training data, with feature randomness introduced at each split to reduce overfitting. The stacking ensemble incorporates multiple base models $(f_1, f_2, \ldots, f_k)$. These models generate predictions on the training set (X_{train}) using out-of-fold predictions to avoid overfitting. Let the base models' predictions be denoted by $\widehat{y_1}, \widehat{y_2}, \ldots, \widehat{y_k}$.

These predictions are concatenated to form a new meta-feature matrix $(X_{\text{meta}} = [\widehat{y_1}, \widehat{y_2}, \ldots, \widehat{y_k}])$. A meta-model (f_{meta}) is then trained on (X_{meta}) to generate the final prediction. Linear Regression is used as the meta-model, which minimizes the residual sum of squares $\min_{\beta_0,\beta_1,\ldots,\beta_k} \sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \sum_{j=1}^{k}\beta_j \widehat{y_{ij}}\right)\right)^2$ where (β_0) and (β_j) are the coefficients learned during the training of the meta-model. GridSearchCV is employed to optimize hyperparameters of each base model, ensuring that the stacking ensemble is fine-tuned for maximum predictive performance.

Model Evaluation

Model evaluation is conducted using a 10-fold cross-validation scheme to ensure robustness and generalizability. Let $(\mathcal{D})$be the dataset, partitioned into 10 subsets. In each iteration, one subset is used as the validation set (V) and the remaining nine subsets as the training set (T). The evaluation metrics computed is presented in equation 5-8.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}| \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{6}$$

$$\text{MSLE} = \frac{1}{n}\sum_{i=1}^{n}(\log(1 + y_i) - \log(1 + \widehat{y_i}))^2 \tag{7}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \times 100 \tag{8}$$

These metrics provide a comprehensive evaluation of model performance, capturing both the accuracy and reliability of the predictions. The final Stacking Regressor model is selected based on its ability to minimize error metrics while maximizing (R^2), demonstrating its superiority over individual baseline models.

RESULT

As presented in the table 1 and figure 2, the results of this study were obtained by evaluating the predictive performance of various regression models, including Linear Regression, Random Forest, Gradient Boosting, XGBoost, LightGBM, Support Vector Regressor, and the Stacking Regressor. The evaluation metrics used were Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared, Mean Squared Log Error (MSLE), and Mean Absolute Percentage Error (MAPE). MAE provides an average of the absolute errors between predicted and actual values, offering an intuitive measure of prediction accuracy. The results indicate that the Support Vector Regressor (MAE: 4.3091 ± 0.2643) and Linear Regression (MAE: 4.3154 ± 0.2676) models performed the best in terms of minimizing the average prediction error. In contrast, the Random Forest model exhibited the highest MAE (4.8558 ± 0.2620), suggesting that it struggled to consistently predict the target variable accurately. Interestingly, the Stacking Regressor (MAE: 4.5340 ± 0.2454) did not outperform the individual base models as expected. While it achieved a relatively low MAE, it was still slightly higher than the Support Vector Regressor and Linear Regression, indicating that the stacking approach might not have fully leveraged the strengths of the base models in this context.

RMSE penalizes larger errors more than MAE, providing insights into the variance of prediction errors. The Support Vector Regressor (RMSE: 5.4110 ± 0.3142) and Linear Regression (RMSE: 5.4137 ± 0.3189) again emerged as the top performers, indicating their robustness in minimizing large deviations in predictions. On the other hand, the Random Forest model recorded the highest RMSE (6.0518 ± 0.3260), further reinforcing the observation that it was less

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, Number 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

effective in capturing the underlying patterns in the data. The Stacking Regressor (RMSE: 5.6762 ± 0.2735) demonstrated a moderate performance, with an RMSE slightly higher than the best individual models but lower than other ensemble methods like Gradient Boosting and XGBoost. This outcome suggests that while the stacking ensemble could integrate the predictions of multiple models, It may have been constrained by the potential limitations of its base learners.

The R-squared metric indicates the proportion of variance in the target variable explained by the model. Higher R-squared values imply a better fit to the data. The Linear Regression (R^2: 0.8685 ± 0.0229) and Support Vector Regressor (R^2: 0.8685 ± 0.0236) achieved the highest values, indicating that they were most effective in explaining the variability in the target variable. Notably, the Stacking Regressor (R^2: 0.8563 ± 0.0203) achieved an R-squared close to the top-performing models, suggesting that it could capture a substantial portion of the variance in the target variable, though not to the same extent as the simpler models. The Random Forest model (R^2: 0.8357 ± 0.0277) had the lowest R-squared, implying that its predictions were less aligned with the actual values, possibly due to overfitting or an inability to generalize effectively on this specific dataset.

MSLE measures the ratio between the predicted and actual values, offering a logarithmic perspective that is particularly useful when dealing with exponential growth. In this analysis, Linear Regression (MSLE: 0.0132 ± 0.0141) and Support Vector Regressor (MSLE: 0.0140 ± 0.0164) once again delivered the lowest error rates, indicating their consistent performance across different error metrics. The ensemble models, including the Stacking Regressor (MSLE: 0.0171 ± 0.0229), demonstrated slightly higher MSLE values, reflecting their varying ability to handle the distribution of errors effectively. Interestingly, Random Forest (MSLE: 0.0191 ± 0.0236) and LightGBM (MSLE: 0.0212 ± 0.0326) recorded the highest MSLE values, suggesting that they struggled with predicting low values accurately. MAPE provides a percentage-based error metric, making it useful for comparing models across different scales. However, the MAPE values in this study were notably high and inconsistent, with Linear Regression showing a MAPE of 34300003477753.4688 ± 102900010433260.2188, and other models displaying similarly large and variable values. These inflated MAPE results indicate that the data may contain extreme values or outliers, which significantly influenced the percentage-based error. Notably, the Stacking Regressor recorded a MAPE of 63569311868352.0391 ± 190707935605055.8750, indicating that while it performed well on other metrics, it struggled to provide accurate percentage-based predictions. The high variability across models for this metric suggests that MAPE may not be the most reliable metric for this dataset due to its sensitivity to extreme values.

The analysis reveals several key insights into the effectiveness of different regression models in predicting student performance. Support Vector Regressor and Linear Regression consistently emerged as the top performers across most metrics, achieving low error rates and high R-squared values. Their simplicity and inherent ability to model linear and non-linear relationships may have contributed to their effectiveness in capturing the underlying patterns in the dataset. This observation challenges the conventional expectation that more complex ensemble methods like the Stacking Regressor would necessarily outperform simpler models. The Stacking Regressor, while designed to leverage the strengths of multiple models, did not significantly outperform the individual base models. Its performance indicates that the potential improvement from combining multiple models might be marginal, particularly when the individual models already exhibit high predictive accuracy. This result underscores the importance of careful model selection and ensemble configuration, as adding complexity does not always translate into better performance.

Table 1. Comparison Results of the Evaluation Metrics

Model	MAE (± Std Dev)	RMSE (± Std Dev)	R^2 (± Std Dev)	MSLE (± Std Dev)	MAPE (± Std Dev)
Linear Regression	4.3154 ± 0.2676	5.4137 ± 0.3189	0.8685 ± 0.0229	0.0132 ± 0.0141	34300003477753.4688 ± 102900010433260.2188
Random Forest	4.8558 ± 0.2620	6.0518 ± 0.3260	0.8357 ± 0.0277	0.0191 ± 0.0236	71261958103759.2344 ± 213785874311277.4375
Gradient Boosting	4.5189 ± 0.2443	5.6572 ± 0.2644	0.8578 ± 0.0197	0.0170 ± 0.0230	64040913252021.2969 ± 192122739756063.6875
XGBoost	4.5088 ± 0.2409	5.6357 ± 0.2789	0.8587 ± 0.0173	0.0174 ± 0.0250	70849535701680.2188 ± 212548607105040.3750
LightGBM	4.7628 ± 0.2628	5.9837 ± 0.3334	0.8404 ± 0.0228	0.0212 ± 0.0326	112178057696788.0156 ± 336534173090363.8125
Support Vector Regressor	4.3091 ± 0.2643	5.4110 ± 0.3142	0.8685 ± 0.0236	0.0140 ± 0.0164	41068602654633.8281 ±

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

Stacking Regressor	4.5340 ± 0.2454	5.6762 ± 0.2735	0.8563 ± 0.0203	0.0171 ± 0.0229	123205807963901.2969 63569311868352.0391 ± 190707935605055.8750

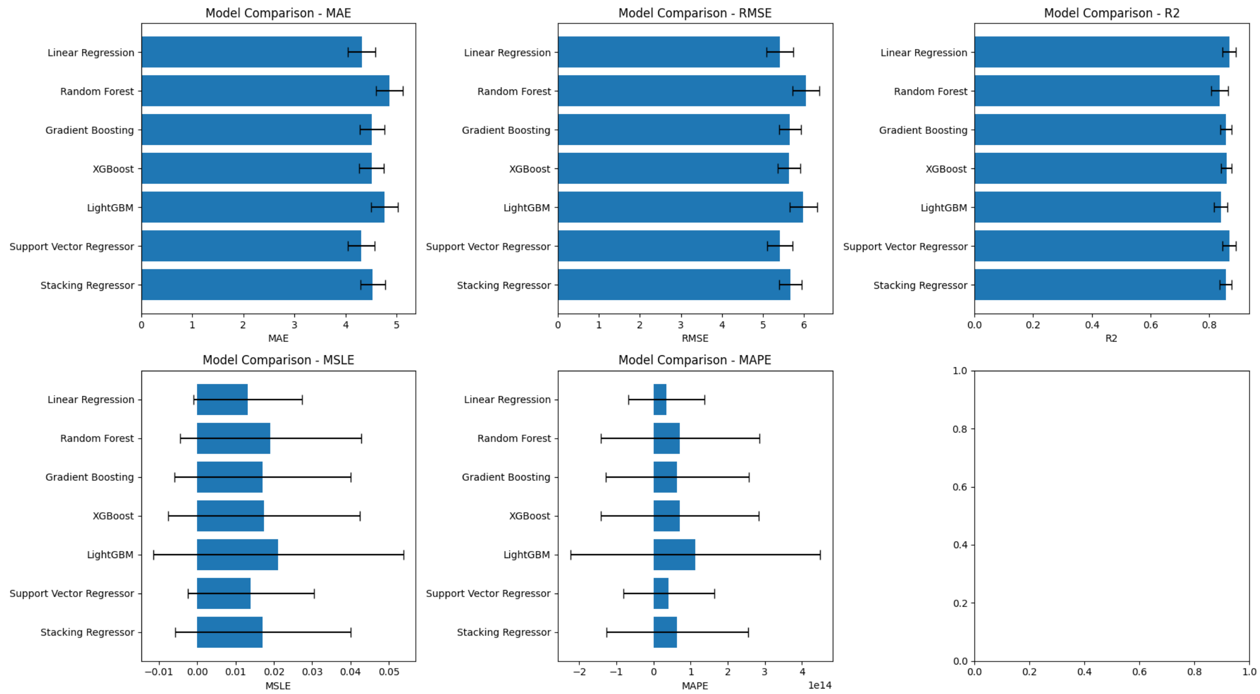

Figure 1. Visualization of each metrics result

CONCLUSION

This study aimed to predict student performance using advanced machine learning models, including traditional regression models, ensemble methods, and a stacking regressor. By evaluating multiple metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared, Mean Squared Log Error (MSLE), and Mean Absolute Percentage Error (MAPE), we aimed to identify the most effective model for capturing the complex relationships between socio-economic, demographic, and academic factors affecting student outcomes. The results provide several key insights into the performance of these models and the dynamics of student performance prediction.

Linear Regression and Support Vector Regressor consistently emerged as the top performers across most evaluation metrics, achieving the lowest errors and highest R-squared values. These models demonstrated robust predictive power, suggesting that despite the complexity of the data, simpler models can effectively capture the key patterns and relationships. This finding challenges the assumption that more sophisticated ensemble methods necessarily outperform simpler models in every context. The stacking regressor, while designed to combine the strengths of multiple models, did not significantly surpass the performance of its individual base models. This suggests that the potential benefits of stacking may be marginal when the individual models already exhibit strong predictive capabilities. The study underscores the importance of careful model selection, emphasizing that added complexity does not always lead to improved performance.

The analysis of MAPE revealed notable variability and inflated values across all models, indicating the presence of extreme values or outliers in the dataset. This finding highlights the need for robust preprocessing techniques to handle outliers more effectively, as they can significantly impact certain evaluation metrics. The inconsistencies in MAPE also suggest that some metrics may be more sensitive to the characteristics of the dataset, thereby emphasizing the importance of using multiple evaluation criteria to obtain a comprehensive assessment of model performance. In conclusion, this research demonstrates that while advanced machine learning techniques like stacking ensembles have the potential to enhance predictive accuracy, simpler models such as Linear Regression and Support Vector Regressor may be sufficient for certain datasets and contexts. The results highlight the necessity for a balanced approach to model selection, taking into account the complexity of the model, interpretability, and the nature of the dataset. Future research could focus on exploring more advanced ensembling strategies, incorporating additional feature engineering, and addressing outliers through more sophisticated data preprocessing methods. Moreover, the findings suggest potential avenues for applying these models in educational policy and intervention, aiding educators in identifying at-risk students and enhancing educational outcomes through data-driven strategies.

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

REFERENCES

Alawsi, M. A., Zubaidi, S. L., Al-Bdairi, N. S. S., Al-Ansari, N. & Hashim, K. (2022). Drought forecasting: a review and assessment of the hybrid techniques and data pre-processing. Hydrology, 9(7), 115.

Badal, Y. T. & Sungkur, R. K. (2023). Predictive modelling and analytics of students' grades using machine learning algorithms. Education and Information Technologies, 28(3), 3027–3057.

Bento, P. M. R., Pombo, J. A. N., Calado, M. R. A. & Mariano, S. J. P. S. (2021). Stacking ensemble methodology using deep learning and ARIMA models for short-term load forecasting. Energies, 14(21), 7378.

Blanden, J. (2020). Education and inequality. In The economics of education (pp. 119–131). Elsevier.

Cohen, J. F. W., Hecht, A. A., McLoughlin, G. M., Turner, L. & Schwartz, M. B. (2021). Universal school meals and associations with student participation, attendance, academic performance, diet quality, food security, and body mass index: a systematic review. Nutrients, 13(3), 911.

D'iez, F., Villa, A., López, A. L. & Iraurgi, I. (2020). Impact of quality management systems in the performance of educational centers: educational policies and management processes. Heliyon, 6(4).

Davids, R. (2020). The nature of parental involvement in literacy activities of low achieving learners in disadvantaged contexts at a selected primary school in the Western Cape.

Domladovac, M. (2021). Comparison of neural network with gradient boosted trees, random forest, logistic regression and SVM in predicting student achievement. 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), 211–216.

Goshin, M., Dubrov, D., Kosaretsky, S. & Grigoryev, D. (2021). The strategies of parental involvement in adolescents' education and extracurricular activities. Journal of Youth and Adolescence, 50(5), 906–920.

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U. & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Applied Sciences, 10(11), 3894.

Hashim, A. S., Awadh, W. A. & Hamoud, A. K. (2020). Student performance prediction model based on supervised machine learning algorithms. IOP Conference Series: Materials Science and Engineering, 928(3), 32019.

Hooda, M., Rana, C., Dahiya, O., Rizwan, A. & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. Mathematical Problems in Engineering, 2022(1), 5215722.

Hultberg, P. T., Calonge, D. S. & Choi, T. (2021). Costs and benefits of private tutoring programs: the South Korean case. International Journal of Social Economics, 48(6), 862–877.

Kabathova, J. & Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques. Applied Sciences, 11(7), 3130.

Ludeke, S. G., Gensowski, M., Junge, S. Y., Kirkpatrick, R. M., John, O. P. & Andersen, S. C. (2021). Does parental education influence child educational outcomes? A developmental analysis in a full-population sample and adoptee design. Journal of Personality and Social Psychology, 120(4), 1074.

Marks, G. N. & O'Connell, M. (2021). Inadequacies in the SES--Achievement model: Evidence from PISA and other studies. Review of Education, 9(3), e3293.

Ofori, F., Maina, E. & Gitonga, R. (2020). Using machine learning algorithms to predict students' performance and improve learning outcome: A literature based review. Journal of Information and Technology, 4(1), 33–55.

Razzaq, H. A. R., Majeed, M. F. M. & Bajwa, M. S. A. B. (2024). Relationship between Socioeconomic Status and Academic Achievements. UCP Journal of Humanities \& Social Sciences (HEC Recognized-Y Category), 2(2), 73–88.

Roberts, M., Tolar-Peterson, T., Reynolds, A., Wall, C., Reeder, N. & Rico Mendez, G. (2022). The effects of nutritional interventions on the cognitive development of preschool-age children: a systematic review. Nutrients, 14(3), 532.

Rodr'iguez-Hernández, C. F., Cascallar, E. & Kyndt, E. (2020). Socio-economic status and academic performance in higher education: A systematic review. Educational Research Review, 29, 100305.

Strelan, P., Osborn, A. & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. Educational Research Review, 30, 100314.

Syed Mustapha, S. (2023). Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods. Applied System Innovation, 6(5), 86.

Tan, C. Y., Lyu, M. & Peng, B. (2020). Academic benefits from parental involvement are stratified by parental socioeconomic status: A meta-analysis. Parenting, 20(4), 241–287.

Villar, A. & de Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study. Discover Artificial Intelligence, 4(1), 2.

Xue, H. & Niu, Y. (2023). Multi-output based hybrid integrated models for student performance prediction. Applied Sciences, 13(9), 5384.

Yaugci, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 11.

E-ISSN : 2807-9035
Volume 4, **Number** 2, November 2024
https://doi.org/10.47709/brilliance.v4i2.4669

Yu, F. & Liu, X. (2022). Research on student performance prediction based on stacking fusion model. Electronics, 11(19), 3166.