Jurnal Statistika Dan Komputasi (STATKOM), Vol. 5 No. 1 (June 2026)
ISSN : 2963-038X (print), ISSN : 2963-0398
DOI : https://doi.org/10.32665/statkom.v5i1.6381

STATKOM
JURNAL STATISTIKA DAN KOMPUTASI

Explainable Gradient Boosting for GPA Prediction with SHAP Aggregation

Jerhi Wahyu Fernanda[1], M Khoiril Akhyar[2], Ninik Zuraidah[3], Novi Rosita Rahmawati[4], Lilla Maturizka Ayu Asfarin[5]

[1,2,3]Educational Mathematical Program, State Islamic University of Syekh Wasil Kediri
[4]Islamic Religious Education Department, State Islamic University of Syekh Wasil Kediri
[5]Guidance and Counseling Department, Universitas Sebelas Maret
E-mail : fernanda.jerhi@uinkediri.ac.id

Submitted: March 13, 2026 ***Revised:*** June 11, 2026 ***Accepted:*** June 16, 2026

Abstract

Background: *The Grade Point Average (GPA) is a complicated quantity impacted by psychological factors as well as the learning environment. Previous research has concentrated exclusively on understanding GPA prediction models through a singular predictive framework, resulting in inconsistent assessments of varied significance.*
Objective: *This study aims to propose a SHAP aggregation framework to explain the importance of features from three gradient boosting models, consisting of XGBoost, LightGBM, and CatBoost.*
Methods: *The dependent variable in this study is student GPA, whereas the predictor variables encompass emotional intelligence, adversity quotient, metacognitive ability, motivation, self-efficacy, and learning environment. The predictor variable was a composite score obtained from the average of the items of each latent variable. Data were collected through a survey of 270 students with probability sampling.*
Results: *The SHAP Importance Feature of the three gradient boosting models exhibited varying outcomes regarding the variables with the greatest and second-highest contributions. XGBoost and LightGBM recognized metacognitive capacity as the predominant attribute, while CatBoost highlighted adversity quotient. The SHAP aggregation results indicate that Adversity Quotient is the predominant factor, with a mean importance feature value of 0.0318.*
Conclusion: *The findings indicate that the SHAP aggregation framework offers a more stable and robust interpretation of variable importance compared to a single model. This methodology can also serve as an analytical framework in educational data mining research.*

Keywords : *Grade Point Average, XGBoost, LightGBM, CatBoost, SHAP Aggregation.*

Abstrak

Latar Belakang: Indeks Prestasi Kumulatif (IPK) adalah variabel kompleks yang dipengaruhi faktor psikologis dan lingkungan belajar. Studi sebelumnya memiliki keterbatasan yaitu fokus kepada interpretasi model prediksi GPA hanya didasarkan pada model prediktif tunggal sehingga berdampak kepada interpretasi yang tidak stabil terkait *importance variable*.
Tujuan: Penelitian ini bertujuan mengusulkan kerangka kerja SHAP *aggregation* untuk menjelaskan *importance* variabel dari tiga model *gradient boosting* yang terdiri dari XGBoost, LightGBM, CatBoost.
Metode: Variabel dependen dalam studi ini adalah IPK mahasiswa, sedangkan variabel prediktor meliputi kecerdasan emosional, *adversity quotient*, kemampuan metakognitif, motivasi, efikasi diri, dan lingkungan belajar. Data dikumpulkan melalui survei dengan total sampel sebanyak 270 mahasiswa.
Hasil: SHAP *Importance Feature* dari 3 model *gradient boosting* menunjukkan hasil yang berbeda pada variabel yang memiliki kontribusi paling besar dan nomor dua. XGBoost dan LightGBM mengidentifikasi kemampuan metakognitif sebagai fitur paling dominan, dan CatBoost mengidentifikasi *adversity quotient*. Hasil SHAP *aggregation* menunjukkan bahwa kemampuan *Adversity Quotient* menjadi faktor terpenting secara keseluruhan dengan nilai mean *importance feature* sebesar 0,0318.
Kesimpulan: Temuan ini menunjukkan bahwa kerangka kerja SHAP *aggregation* memiliki interpretasi yang lebih stabil dan kuat terkait *importance variable* dibandingkan model tunggal. Pendekatan ini juga dapat digunakan untuk kerangka analitis dalam penelitian *Educational Data Mining.*

Kata kunci: Indeks Prestasi Kumulatif (IPK), XGBoost, LightGBM, CatBoost, SHAP *Aggregation*.

INTRODUCTION

The Grade Point Average (GPA) is a complex quantity affected by cognitive intelligence, various other intelligences, and psychological factors. Previous studies have shown that GPA is not solely dependent on intellectual intelligence but is also influenced by emotional intelligence and adversity quotient (Nieto-Carracedo et al., 2024; Rena et al., 2023). Furthermore, psychological factors encompass self-efficacy, motivation, and metacognition, which would affect GPA (Amien, Hidayatullah, 2023). In addition, external factors such as the learning environment will have a positive impact on achieving better academic outcomes (Lu et al., 2024). The complexity of predictor variables requires analytical methods that offer both high precision and significant interpretability, facilitating a clear and transparent evaluation of each variable's contribution.

Explainable Artificial Intelligence (XAI) is a methodology for developing AI models, including Deep Learning, Random Forest, and XGBoost, to enhance the reliability and transparency of the prediction for human users (Kabir & Hossain, 2025). SHapley Additive Explanation (SHAP) is an XAI method based on game theory that serves as the inspiration for this technique and illustrates how each feature influences the prediction (Ali et al., 2023). Previous research has used this method to understand the contribution of each feature in a machine learning model (Choi et al., 2024; Fernanda et al., 2025; Santos and Guedes, 2024).

Moreover, studies utilizing XAI and SHAP to predict GPA or in education were limited. Earlier studies on GPA can be classified into two groups. The first group emphasizes using statistical analysis methods like variance-based and covariance-based structural equation models (Falebita & Asanre, 2025; Herlina et al., 2025). The second group was working on machine learning research to find the best model for predicting GPA by testing different machine learning methods, including research that uses the average composite score of latent variables as predictor variables (Bhushan et al., 2024; Gufroni et al., 2025; Lu et al., 2024). The analysis of previous research concerning GPA reveals a substantial gap for exploration by amalgamating feature importance results from various machine learning models via aggregated SHAP, thus attaining enhanced stability and robustness in feature importance analysis, which could lead to more reliable predictions and insights in academic performance analysis.

In the field of artificial intelligence and machine learning, gradient boosting is a group of techniques that includes Extreme Gradient Boosting (XGBoost), LightGradient Boosting (LightGBM), and Categorical Boosting (CatBoost). These techniques are used for data analytics (Daniel, 2024). Gradient boosting offers advantages in modeling by effectively capturing non-linear correlations and complicated interactions among variables, leading to superior prediction performance compared to traditional techniques (Rizkallah, 2025). These techniques concentrate on iterative modeling that emphasizes mistakes in the preceding model. XGBoost, which optimizes regulation and parallel computing (Su et al., 2023). LightGBM accelerates the training process through leaf-wise growth and histogram-based learning techniques (Ramalingam et al., 2024), and CatBoost was designed to handle categorical or mixed variables more effectively and reduce prediction shift (Geeitha et al., 2024). Although these models had superior performance. There is a lack of interpretation regarding the model's prediction. Consequently, incorporating interpretability approaches like SHAP was essential for quantitatively elucidating the contribution of each feature to the prediction.

The objective of this study is to determine the importance of features in predicting GPA through SHAP aggregation applied to the XGBoost, LightGBM, and CatBoost models. The findings of this study will provide educators with strategies to improve student academic performance through psychological factors or to enhance the learning environment. This study provides a framework for a more stable and resilient interpretable machine learning model.

METHOD

Research Design

This study used a cross-sectional methodology to assess GPA and the variables that influence it.

Population and Sample

This study's population comprised 3.382 students from the Faculty of Education and Teacher Training, Islamic State University of Syekh Wasil Kediri. The sampling technique used was probability sampling, with the sample size determined using Formula 1. The p-value of 0.5 is based on the unknown probability of the proportion of students' GPAs below 3.25. This value will give maximum variability and good precision (Ahmed, 2024).

$$n = \frac{z^2_{1-\alpha/2}\, p\,(1-p)N}{e^2(N-1)+z^2_{1-\alpha/2}\, p\,(1-p)} \qquad (1)$$

N = 3382

$\alpha = 0.1$

$z^2_{1-\alpha/2} = 1.645^2$

e = 0.05

p = 0.5

$$n = \frac{1.645^2 0.5\ 0.5\ 3382}{0.05^2(3382)+1.645^2 0.5\ 0.5} = 250.62$$

$$n = 251$$

The actual sample of this study was 253 respondents.

Sampling Techniques

This study used a probability sampling technique to select respondents who met specific inclusion criteria, that was completed at least two semesters, and were included in the study to ensure that all respondents had an established Grade Point Average (GPA). Data were collected using a questionnaire distributed through Google Forms.

Research Subjects

This study uses primary data, including a dependent variable (GPA) and six independent variables: motivation, self-efficacy, metacognition, emotional intelligence, adversity quotient, and learning environment. GPA represents academic performance, while the other variables serve as predictors.

Measurement Instrument Design

This research used a questionnaire to assess predictor variables. Motivation and self-efficacy were measured with an instrument shown to be valid and reliable in a preliminary survey of 30 students. All questions had correlation values above 0.361, and all latent variables had Cronbach's alpha values above 0.6.

Table 1 describes the number of indicators and the data scale for each variable.

Table 1. Description of Variables

Variable	Total of Question	Scale
Motivation	6	Likert scale with a score of 1: Strongly Disagree to 4: Strongly Agree
Self Efficacy	6	Likert scale with a score of 1: Strongly Disagree to 4: Strongly Agree
Metacognitive	10	Likert scale with a score of 1: Strongly Disagree to 4: Strongly Agree
Emotional Intellegent	10	Likert scale with a score of 1: Strongly Disagree to 4: Strongly Agree

Variable	Total of Question	Scale
Adversity Quotient	24	Likert scale with a score of 1: Strongly Disagree to 4: Strongly Agree
learning environment	7	Likert scale with a score of 1: Strongly Disagree to 4: Strongly Agree
GPA	-	Ratio

Data Analysis Techniques

The SHAP Aggregation Analysis Framework Using XGBoost, LightGBM, and CatBoost consists of 6 stages with the following explanation.

1. The research data, comprising GPA, motivation, self-efficacy, metacognition, emotional intelligence, adversity quotient, and learning environment, is partitioned into training and testing datasets in an 80:20 ratio in the training process. The data validation process is carried out using 10-fold cross-validation. This procedure provides an optimal balance in the data training process and increases its effectiveness (Malakouti et al., 2023; Sivakumar & Parthasarathy, 2024).
2. Normalize data with min-max scaler. Normalization in stage 2 was performed by transforming the predictor data into a range of 0-1 with the formula. This process will improve model performance (Kim et al., 2025).
3. Data modeling and validation using XGBoost, LightGBM, and CatBoost on training data and validation with 10-fold cross-validation. Xgboost, LightGBM, and CatBoost are ensemble learning models within the gradient boosting framework that work by iterating residuals from the previous stage as targets for the next iteration (Su et al., 2023). XGBoost, LightGBM, and Catboost optimize the following objective function.

$$L^{(t)} = \sum_{i=1}^{n}\left[l\left(y_i, y_i^{(t-1)}\right) + g_i f_i(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) \qquad (2)$$

wheres

$L^{(t)}$: objective function in t iteration
n : sample
y_i : actual data
$y_i^{(t-1)}$: predicted values from previous iteration.
x_i : i-th observation predictor
$f_i(x_i)$: new tree added at iteration t
$l\left(y_i, y_i^{(t-1)}\right)$: Loss function that measures the prediction error.
g_i: gradient
h_i : hessian
$\Omega(f_t)$: regularization function.

Although all three methods have the same objective function, the differences lie in the gradient estimation mechanism, split strategy, and tree growth structure. XGBoost optimizes the objective function using a second-order gradient approach and level-wise tree growth.

LightGBM uses histogram-based splitting, Gradient-based One-Side Sampling (GOSS), Exclusive Feature Bundling (EFB), and leaf-wise tree growth, which impacts speed and efficiency when working with big data. Leaf-wise Tree Growth is a procedure that selects nodes with maximum delta losses during tree construction. GOSS works by selecting features that have the largest gradient and focusing on samples that have the largest contribution to the prediction error. Exclusive Feature Bundling reduces the number of features by grouping features that have mutually exclusive value into one feature (Ramalingam et al., 2024).

CatBoost works with an Ordered Boosting mechanism and tree growth with a symmetric tree (Geeitha et al., 2024).

$$g_i^{ordered} = \frac{\partial \iota\left(y_i, \hat{y}_i^{ordered}\right)}{\partial \hat{y}_i^{ordere}}$$

$\hat{y}_i^{ordered} = F(X_I|\sigma_1, \dots, \sigma_{1-1})$: prediction for the i-th observation uses only the previous observations in a data permutation

4. Calculate the RMSE value of XGBoost, LightGBM, and CatBoost from testing data using the formula.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (3)$$

Inverse RMSE was used to calculate inverse RMSE (Chen & Zheng, 2025).

$$W_i = \frac{1/RMSE_i}{\sum_{i=1}^{3}(1/RMSE_j)} \quad (4)$$

$0 < W_i < 1, \sum_1^3 W_i = 1$

W_i was weighted RMSE value of each model (XGBoost, LightGBM, and CatBoost).

5. Calculate the SHAP aggregation of each feature with formula

$$\begin{aligned} SHAP\ Aggregation\ (f) &= W_{XGBoost} SHAP_{XGBoost}(f) \\ &+ W_{LightGBM} SHAP_{LightGBM}(f) \\ &+ W_{CatBoost} SHAP_{CatBoost}(f) \end{aligned} \quad (5)$$

$SHAP_i(f)$ is the contribution of feature to i model = XGBoost, LightGBM, Catboost .

6. Identification of variables with the most significant impact based on the SHAP Aggregation.

The comprehensive data analysis in this study employed Jupyter Notebook and Python libraries. Figure 1 additionally elucidates the data analysis framework.

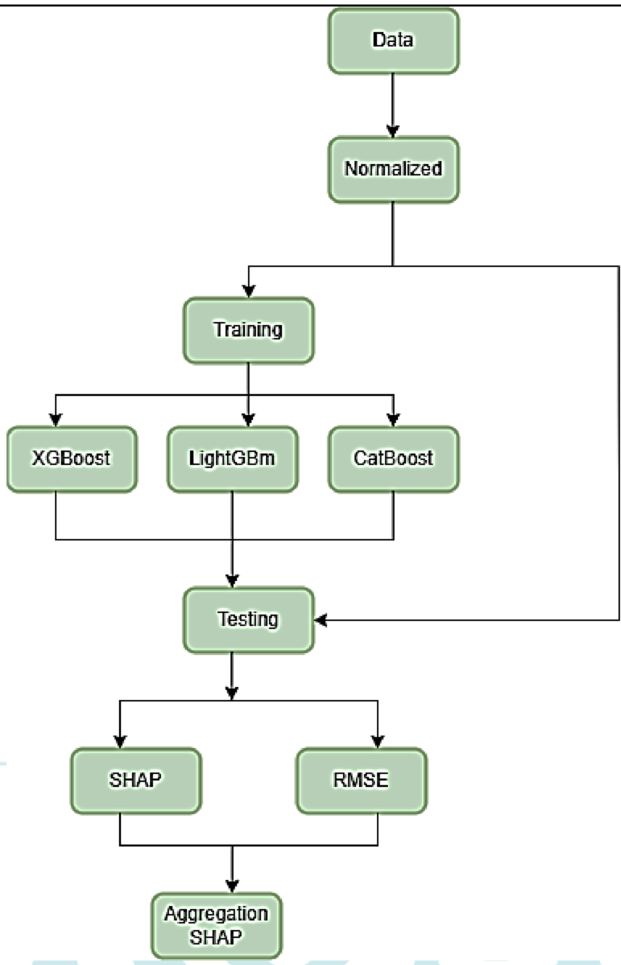


Figure 1. Data Analysis flow chart

RESULTS AND DISCUSSION

The GPA distribution in the Faculty of Education and Teacher Training at the Islamic State University of Syekh Wasil Kediri was left-skewed pattern. This pattern indicates excellent student academic achievement. The majority of students have GPAs above 3.5. These results also indicate that the learning process has been running optimally.

Motivation has a distribution pattern approaching a normal distribution, centered around its respective mean values. The environment has a data pattern heavily concentrated at high values. Metacognitive and Emotional Quotient variables have a bimodal pattern, and even Adversity Quotient shows this pattern clearly. This bimodal pattern provides insight that these variables can be grouped into two groups due to data heterogeneity.

The Heatmap correlation explained the

association among variables. This heatmap is also utilized to identify multicollinearity between independent variables. The metacognitive variable exhibits the strongest association with GPA among the independent variables, with a correlation coefficient of 0.33. The heatmap indicates significant relationships among independent variables, notably between the Emotional Quotient and Environmental variables, with a correlation coefficient surpassing 0.5.

The various data patterns shown in Figure 1, combined with the heatmap correlation results that show multicollinearity and weak linear correlation between independent and dependent variables, established a foundation to use the XGBoost method for data analytics. Performance for XGBoost, LightGBM, and CatBoost was measured with RMSE in Table 2.

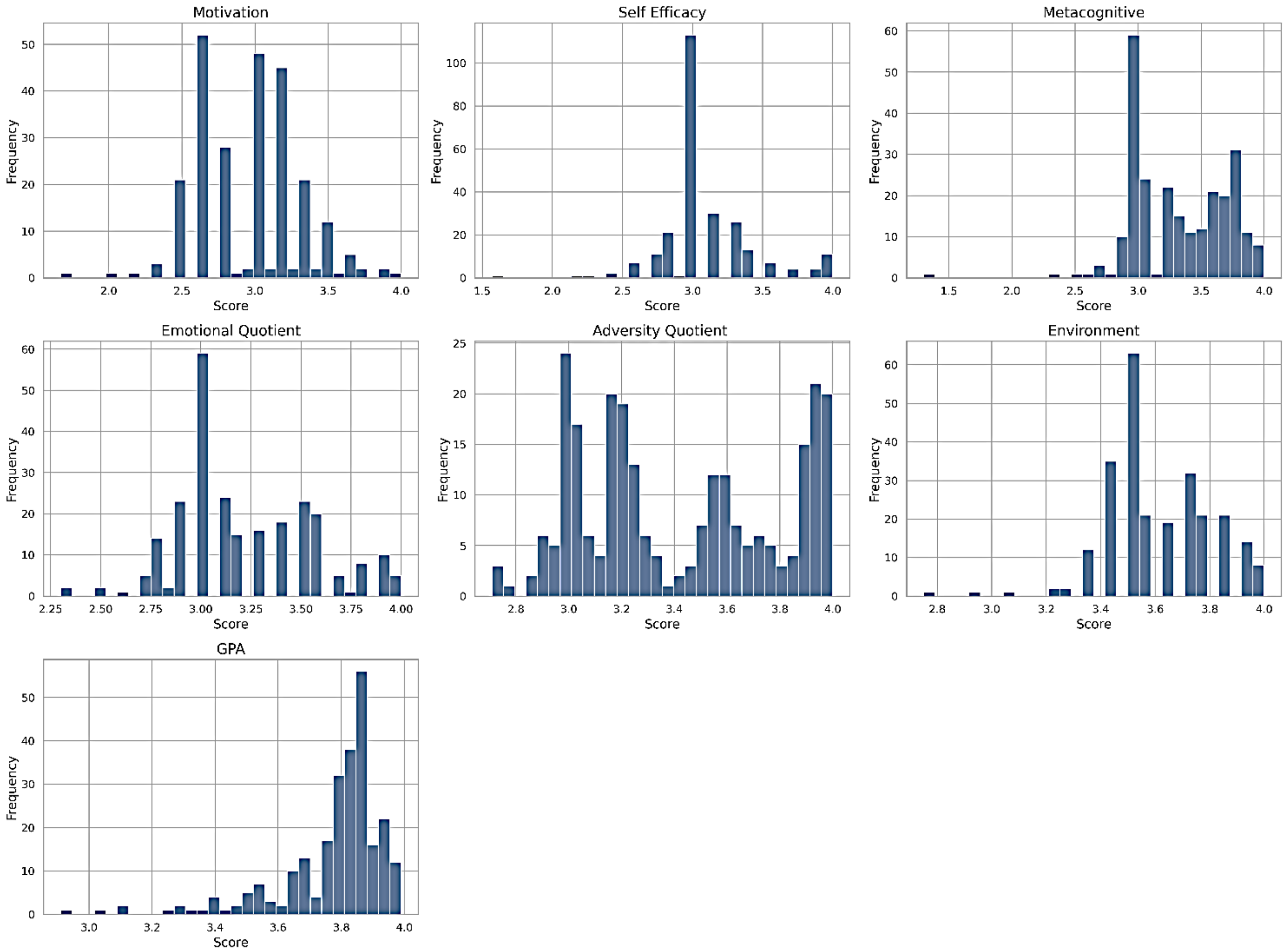


Figure 2. Histogram of variables

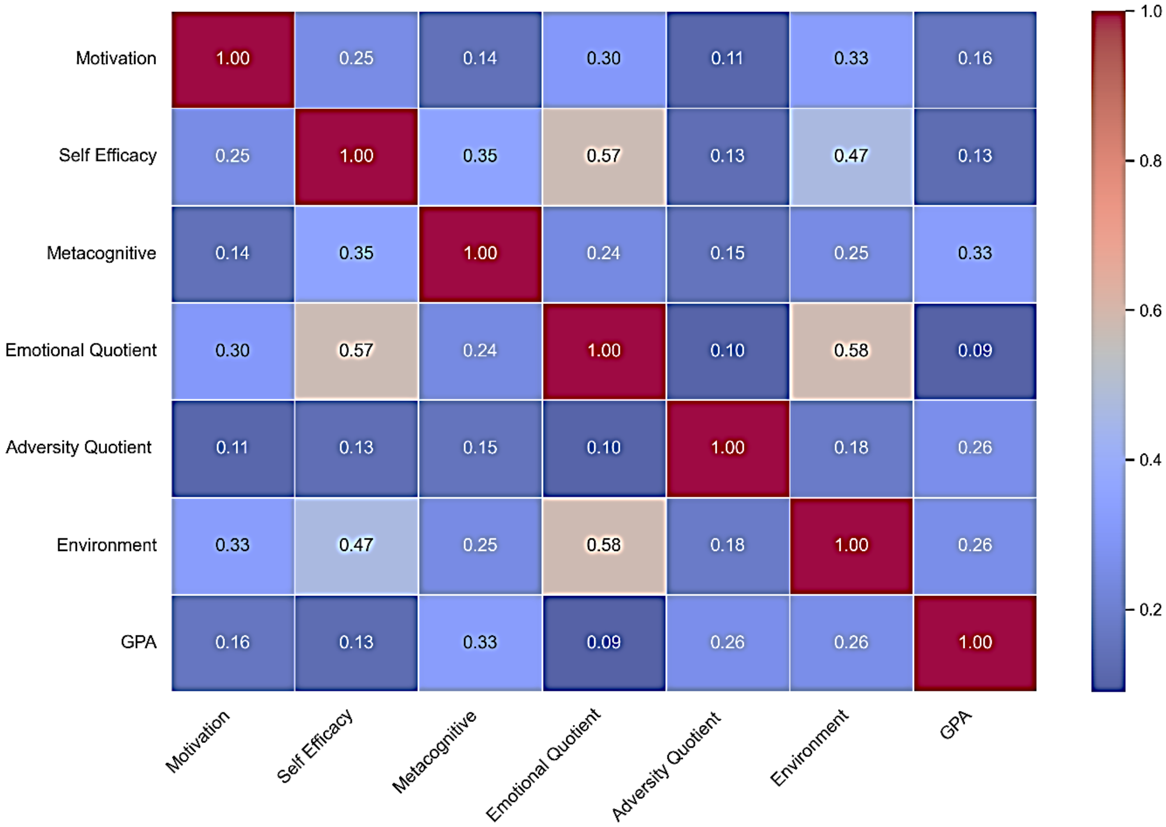
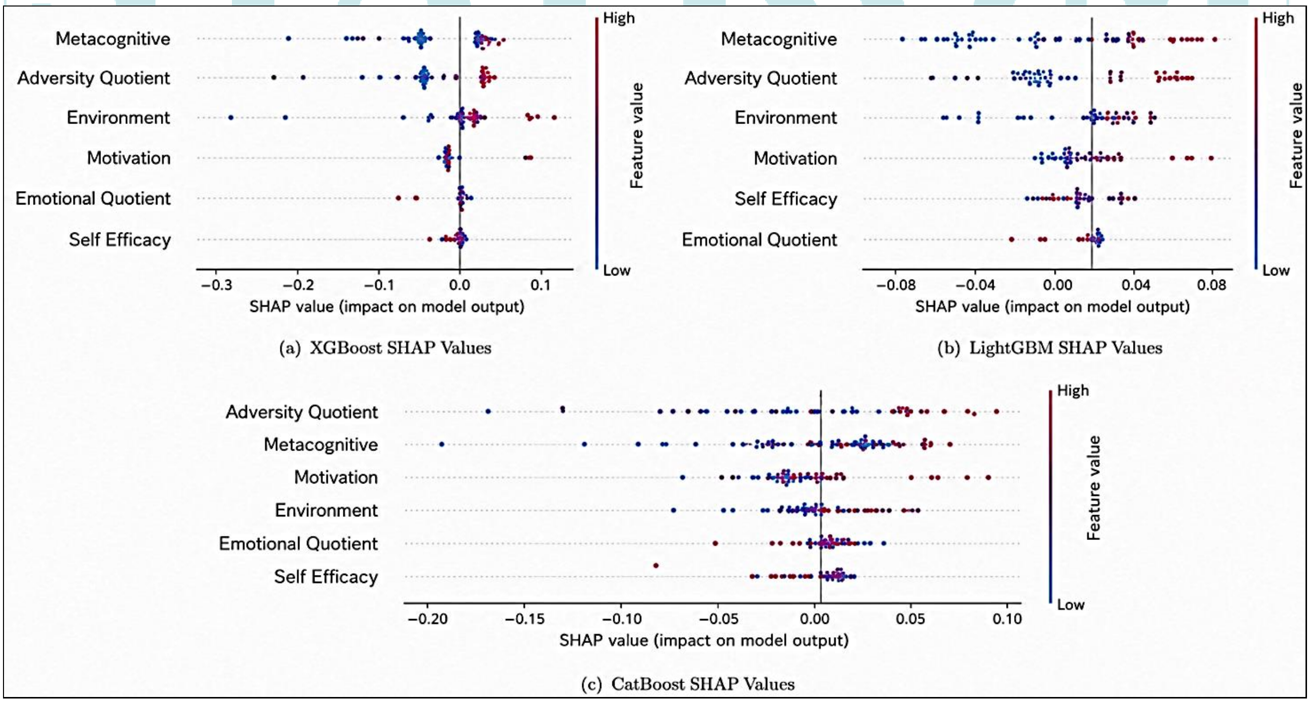


Figure 3. Correlation Heatmap



(a) XGBoost SHAP Values

(b) LightGBM SHAP Values

(c) CatBoost SHAP Values

Figure 4. SHAP Values

Figure 5. SHAP Mean Importance Feature

Table 2. RMSE and Weighted Importance Feature

Methods	RMSE	Weighted SHAP
XGBoost	0.1141	0.386
LightGBM	0.153	0.288
CatBoost	0.1353	0.326

XGBoost yields the lowest RMSE of 0.1141, signifying that this model delivers the GPA prediction nearest to the actual value. CatBoost ranks second with an RMSE of 0.1353, while LightGBM follows with an RMSE of 0.153, which is the model with the least predictive accuracy among the three approaches. An RMSE of 0.1141 indicates that the mean prediction error for GPA is approximately 0.11 points, representing a minimal error rate on a scale of 1–4. This signifies that XGBoost was superior to other gradient boosting machines.

The model weight calculation, utilizing the inverse RMSE weighting method, indicates that XGBoost achieved the highest weight of 0.3863, followed by CatBoost at 0.3256, and LightGBM at 0.2881. The weights indicate the proportional contribution of each model to the aggregation process, with models exhibiting lower prediction error rates having a greater contribution.

This weight distribution guarantees that the model aggregation process proportionately represents the predictive quality of each algorithm, leading to a more robust and representative estimation of feature contributions within the SHAP Aggregation. Previous machine learning studies have only focused on determining the contribution of each feature based on a single model.

Figures 4(a),(b), and (c) present SHAP summary graphs illustrating the distribution of contributions from each predictor variable to GPA prediction across the three models. Each point denotes a single observation, with its position on the horizontal axis signifying the SHAP value that illustrates the extent of the variable's impact on the model output. Red signifies elevated feature values, whereas blue denotes diminished values. In the XGBoost model, the Metacognitive variable has the most extensive range of contributions relative to other factors, suggesting that alterations in metacognitive values significantly impact differences in GPA prediction. In the LightGBM model, the SHAP distribution pattern exhibits a very stable trend regarding contributions to the Metacognitive and Adversity Quotient variables. In Contrast, the CatBoost model had a different result. The Adversity Quotient variable exhibits the most pronounced SHAP value distribution relative to other variables. Nonetheless, this visual representation remains distributional; so, to achieve a better quantitative understanding of each variable's contribution, an analysis of the mean SHAP values in Figure 5 is conducted.

Figure 5 indicates that the metacognitive variable had the greatest XGBoost mean SHAP value of 0.039, establishing it as the primary factor in GPA prediction with a coefficient of 0.039. This number indicates an average increase of 0.039 points in estimated GPA for each unit rise in metacognitive score. The significance of this coefficient indicates that students' capacity to regulate, monitor, and assess their cognitive processes significantly influences disparities in academic performance. Consequently, enhancing metacognitive abilities can substantially influence the improvement of overall academic performance.

The Adversity Quotient (AQ) from the XGBoost Model had a mean SHAP value of 0.025, signifying that each one-unit increment in AQ correlates with an average alteration of 0.025 points in predicted GPA. Despite its lesser contribution compared to metacognition, this value nonetheless demonstrates the crucial role of psychological resilience in affecting academic performance. An elevated AQ enables students to better adjust to academic pressure, learning challenges, and other demanding circumstances, hence positively impacting academic performance, but with mild intensity.

The mean SHAP value for the LightGBM model utilizing CatBoost is illustrated in Figures 5 (b) and (c). In the LightGBM model, the variable with the highest contribution was the metacognitive variable, exhibiting a mean SHAP value of 0.036. The variable contribution levels in LightGBM have a consistent order with those in the XGBoost model. Different results appear in the CatBoost model. In this model. Adversity Quotient was the variable that contributed most significantly to GPA prediction. SHAP aggregation has been employed to integrate the mean importance feature from three models, yielding more consistent results.

SHAP aggregation has different results compared to XGBoost and LightGBM models. The highest contribution variable at SHAP aggregation was Adversity Quotient. This disparity arises from a weighting technique predicated on the Root Mean Square Error (RMSE) value of each model. This method employs an inverse RMSE weighting technique for the SHAP contributions from the XGBoost, LightGBM, and CatBoost models, whereby models with smaller prediction errors are assigned greater weight, rather than merely combining them. The computation of model weights is executed using formula (2).

According to the acquired RMSE values, XGBoost is assigned the highest weight in the aggregation. This weighting technique represents the idea that models with better prediction performance improve the combination of SHAP interpretations across models. Weighted RMSE also has an impact on reducing the mean importance value of SHAP features, which is lower than that of the XGBoost model. This also verifies that SHAP aggregation quantifies feature contributions via consensus, rather than relying on the performance of an individual model, hence yielding more consistent findings.

This SHAP aggregation ensures that the final interpretation is not biased towards one algorithm, but rather reflects a combination of all models that is more robust and stable in explaining the contribution of psychological and environmental elements to GPA prediction. The SHAP aggregation analysis's findings give educators insight into how increasing a student's Adversity Quotient can affect their GPA. This result is consistent with studies that discovered the Adversity Quotient to be an important factor in raising a student's GPA (Safi'i et al., 2021; Wang et al., 2025).

CONCLUSION

Conclusions

This model yields findings for the contribution of each feature that is more moderate yet methodologically more stable and robust, as it integrates a linear combination of the XGBoost, LightGBM, and CatBoost models rather than relying on a singular model.

Suggestions

Recommendations for future study involve the creation of a meta-learning-based ensemble model to forecast student GPA. This approach is able to provide predictions from various machine learning models and has great potential to increase accuracy compared to using conventional weighted averages. Moreover, the research could be augmented by including additional psychological characteristics, such as subjective well-being and resilience, and by extending the group to include university students. These variables are theoretically closely related to academic achievement and are able to improve the interpretation of the predictive model. This methodology can also serve as an analytical framework in educational data mining research, which currently focuses on selecting a single model for interpreting the importance of variables.

ACKNOWLEDGMENTS

The authors highly appreciate the financial support provided for this research. This study was funded by the 2025 Universitas Islam Negeri Syekh Wasil Kediri interdisciplinary basic research cluster grant. The authors also wish to thank the institution for providing the resources and administrative support that made this project possible.

BIBLIOGRAPHY

Ahmed, S. K. (2024). How to choose a sampling technique and determine sample size for research: A simplified guide for researchers. *Oral Oncology Reports*, 12, 100662. https://doi.org/10.1016/j.oor.2024.100662

Ali, S., Akhlaq, F., Imran, A. S., Kastrati, Z., Daudpota, S. M., & Moosa, M. (2023). The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. *Computers in Biology and Medicine*, 166, 107555. https://doi.org/10.1016/j.compbiomed.2023.107555

Amien, M. S., & Hidayatullah, A. (2023). Assessing students' metacognitive strategies in e-learning and their role in academic performance. *Jurnal Inovasi Teknologi Pendidikan*, 10(2), 158–166. https://doi.org/10.21831/jitp.v10i2.60949

Bhushan, M., Verma, U., Garg, C., & Negi, A. (2024). Machine Learning-Based Academic Result Prediction System. *International Journal of Software Innovation*, 12(1). https://doi.org/https://doi.org/10.4018/IJSI.334715

Chen, S., & Zheng, W. (2025). RRMSE-enhanced weighted voting regressor for improved ensemble regression. *PLoS ONE*, 20(3), e0319515. https://doi.org/10.1371/journal.pone.0319515

Choi, J., Choi, Y., Lee, K., Ahn, K., & Jang, W. Y. (2024). Explainable Model Using Shapley Additive Explanations Approach on Wound Infection after Wide Soft Tissue Sarcoma Resection: "Big Data" Analysis Based on Health Insurance Review and Assessment Service Hub. *Medicina*, 60(2), 327. https://doi.org/10.3390/medicina60020327

Daniel, C. (2024). Heliyon A robust LightGBM model for concrete tensile strength forecast to aid in resilience-based structure strategies. *Heliyon*, 10(20), e39679. https://doi.org/10.1016/j.heliyon.2024.e39679

Falebita, O. S., Asanre, A. A., & Chibisa, A. (2025). A structural equation modeling of predictive factors of mathematics undergraduates academic achievement. *Frontiers in Education*, 10. https://doi.org/10.3389/feduc.2025.1572840

Fernanda, J. W., Tsani, I., & Nuraini, A. (2025). Interpretable Ensemble Learning for Online Public Acces Catalog Technology Acceptance Prediction. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, 9(3), 961. https://doi.org/10.31764/jtam.v9i3.30262

Geeitha, S., Ravishankar, K., Cho, J., & Easwaramoorthy, S. V. (2024). Integrating cat boost algorithm with triangulating feature importance to predict survival outcome in recurrent cervical cancer. *Scientific Reports*, 14(1), 1–19. https://doi.org/10.1038/s41598-024-67562-0

Gufroni, A. I., Purwanto, P., & Farikhin, F. (2025). Academic performance prediction using supervised learning algorithms in university admission. *JOIV International Journal on Informatics Visualization*, 9(1), 184. https://doi.org/10.62527/joiv.9.1.2974

Herlina, M., Azizah, N., Rifai, K., Agustin, D., & Sirodj, N. (2025). Modeling the Satisfaction of Data Literacy Online Training for High School Teachers Using PLS SEM. *Jurnal Statistika Dan Komputasi (STATKOM)*, 4(1), 23–32. https://doi.org/10.32665/statkom.v4i1.4429

Kabir, S., Hossain, M. S., & Andersson, K. (2025). A Review of Explainable Artificial Intelligence from the Perspectives of Challenges and Opportunities. *Algorithms*, 18(9), 556. https://doi.org/10.3390/a18090556

Kim, Y., Kim, M. K., Fu, N., Liu, J., Wang, J., & Srebric, J. (2024). Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models. *Sustainable Cities and Society*, 118, 105570. https://doi.org/10.1016/j.scs.2024.105570

Lu, J., Liu, Y., Liu, S., Yan, Z., Zhao, X., Zhang, Y., Yang, C., Zhang, H., Su, W., & Zhao, P. (2024). Machine learning analysis of factors affecting college students' academic performance. *Frontiers in Psychology*, 15(December), 1–14. https://doi.org/10.3389/fpsyg.2024.1447825

Malakouti, S. M., Menhaj, M. B., & Suratgar, A. A. (2023). The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. *Cleaner Engineering and Technology*, 15(July), 100664. https://doi.org/10.1016/j.clet.2023.100664

Nieto-Carracedo, A., Gómez-Iñiguez, C., Tamayo, L. A., & Igartua, J. J. (2024). Emotional Intelligence and Academic Achievement Relationship: Emotional Well-being, Motivation, and Learning Strategies as Mediating Factors. *Psicologia Educativa*, 30(1), 67–74. https://doi.org/10.5093/psed2024a7

Ramalingam, K., Yadalam, P. K., Ramani, P., Krishna, M., Hafedh, S., Badnjević, A., Cervino, G., & Minervini, G. (2024). Light gradient boosting-based prediction of quality of life among oral cancer-treated patients. *BMC Oral Health*, 24(1), 1–8. https://doi.org/10.1186/s12903-024-04050-x

Rizkallah, L. W. (2025). Enhancing the performance of gradient boosting trees on regression problems. *Journal of Big Data*. https://doi.org/10.1186/s40537-025-01071-3

Safi'i, A., Muttaqin, I., Sukino, Hamzah, N., Chotimah, C., Junaris, I., & Rifa'i, M. K. (2021). The effect of the adversity quotient on student performance, student learning autonomy and student achievement in the COVID-19 pandemic era: evidence from Indonesia. *Heliyon*, 7(12). https://doi.org/10.1016/j.heliyon.2021.e08510

Santos, M. R., Guedes, A., & Sanchez-Gendriz, I. (2024). SHapley Additive exPlanations (SHAP) for Efficient Feature Selection in Rolling Bearing Fault Diagnosis. Machine Learning and Knowledge Extraction, 6(1), 316-341. https://doi.org/10.3390/make6010016

Sivakumar, M., & Parthasarathy, S. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. 1–17. https://doi.org/10.7717/peerj-cs.2245

Su, W., Jiang, F., Shi, C., Wu, D., Liu, L., Li, S., Yuan, Y., & Shi, J. (2023). An XGBoost-Based Knowledge Tracing Model. *International Journal of Computational Intelligence Systems*, 16(1), 13. https://doi.org/10.1007/s44196-023-00192-y

Rena, S., Nabila, P., Akbar, M., Fhatona, M. I., & Nafisah, B. (2023). THE RELATIONSHIP BETWEEN EMOTIONAL INTELLIGENCE AND STUDENTS' LEARNING ACHIEVEMENT. *Journal of Psychology and Social Sciences*, 1(4), 143–150. https://doi.org/10.61994/jpss.v1i4.9

Wang, X., Yan, Z., Tang, A., Chen, C., Chen, J., & Xiong, Y. (2025). Adversity quotient influences Self-Regulated learning strategies via achievement motivation among Chinese university students. *Education Sciences*, 15(8), 1042. https://doi.org/10.3390/educsci15081042