Predicting Student Academic Performance in Secondary Education Using Machine Learning Techniques

Turino[1,*], Titi Safitri Maharani[2]

[1,2] Informatics Department, Faculty of Science and Technology, Universitas Perwira Purbalingga, Indonesia

ABSTRACT

Predicting student academic performance has become a critical area of educational research, particularly with the increasing availability of large-scale data and the advancement of artificial intelligence applications in learning analytics. This study investigates the effectiveness of machine learning techniques in forecasting academic performance among secondary school students using state-level educational data. Two ensemble-based algorithms, Random Forest and Gradient Boosting, were implemented to model the complex relationships between student, institutional, and demographic variables. The models were evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2) to ensure robust performance assessment. The Gradient Boosting model achieved the highest predictive accuracy, with an R^2 score of approximately 0.68, outperforming the Random Forest model, which achieved an R^2 of 0.58. Feature importance analysis revealed that factors such as exam participation, gender-based pass rates, teacher availability, school infrastructure, and inclusive education indicators were the most significant predictors of academic outcomes. These findings highlight the role of machine learning as a powerful tool for educational data mining, enabling policymakers and educators to identify key determinants of student success and design data-driven interventions. Future work should incorporate socio-economic and behavioral factors and employ explainable AI approaches to improve the transparency, fairness, and interpretability of predictive educational models.

Keywords Machine Learning, Education, Prediction, Student Performance, Artificial Intelligence

Submitted 9 January 2026
Accepted 5 March 2026
Published 1 June 2026

Corresponding author
Turino,
turinoibrahim07@gmail.com

Additional Information and Declarations can be found on page 82

DOI: 10.63913/ail.v2i2.52

Introduction

Education is widely recognized as a cornerstone of social and economic progress, and improving student academic performance remains a global priority for educators, policymakers, and researchers. The ability to understand and predict student performance plays a key role in shaping effective teaching strategies, developing targeted interventions, and optimizing the allocation of educational resources. Traditionally, student achievement has been analyzed using statistical and econometric approaches such as linear regression or correlation analysis. While these methods offer valuable insights, they often fail to capture the complex, nonlinear interactions between the numerous factors that influence learning outcomes [1], [2]. Student performance is rarely determined by a single variable; rather, it emerges from an interplay of academic, demographic, institutional, and socio-economic conditions. As educational datasets become increasingly large and multidimensional, traditional analytic methods are proving inadequate for extracting meaningful and actionable insights from such complexity [3].

The emergence of Artificial Intelligence (AI) and Machine Learning (ML) has

How to cite this article: Turino and T. S. Maharani, "Predicting Student Academic Performance in Secondary Education Using Machine Learning Techniques," *Artif. Intell. Learn.*, vol. 2, no. 2, pp. 73-84, 2026.

opened new possibilities for analyzing educational data with higher precision and scalability. Machine learning techniques, in particular, are capable of identifying subtle patterns and nonlinear dependencies within datasets that are difficult to detect using conventional methods [4]. Recent studies have highlighted the effectiveness of machine learning models such as Decision Trees, Support Vector Machines, and Neural Networks for predicting academic outcomes in higher education [5]. However, these single models often exhibit limited generalization ability when applied to complex educational environments. To address this limitation, researchers have increasingly turned to ensemble learning approaches, which integrate multiple base models to improve prediction accuracy and robustness [6], [7].

Several studies have demonstrated the superiority of ensemble methods over traditional models. For instance, Balcıoğlu and Artar [8] used the Open University Learning Analytics dataset to compare ensemble algorithms with deep learning models and found that ensemble techniques achieved the highest accuracy and recall in predicting student success. Similarly, Wu et al. [9] conducted a systematic review of 83 studies and concluded that ensemble learning approaches consistently outperform traditional ML models, achieving an average accuracy of 87.67%. Ensemble models such as Random Forest (RF), Gradient Boosting (GB), and XGBoost have proven especially effective at handling high-dimensional educational data, detecting nonlinear relationships, and providing interpretable results [10], [11], [12].

Despite these advancements, existing research often focuses on localized datasets or small institutional samples, limiting the scalability of predictive models [13]. Moreover, while ensemble methods are widely applied in online learning and university settings, there is limited research exploring their use in large-scale, state-level education systems. Addressing this gap is crucial for developing equitable and data-driven educational policies, particularly in developing countries where variability in infrastructure and resource allocation significantly affects student performance [14].

Therefore, this study aims to bridge this gap by applying ensemble-based machine learning techniques, specifically Random Forest and Gradient Boosting, to predict student academic performance using comprehensive state-level secondary education data. The main objectives are threefold: (1) to develop and evaluate machine learning models capable of accurately predicting student outcomes based on educational indicators; (2) to identify and interpret the most influential features affecting academic achievement; and (3) to demonstrate how predictive modeling can inform policy decisions and improve educational planning. By combining predictive accuracy with interpretability, this research contributes to the growing field of AI-driven educational analytics and supports the implementation of machine learning in evidence-based educational policy design [15].

Literature Review

Machine learning and artificial intelligence have become powerful tools for predicting student academic performance. Recent systematic reviews have shown that ensemble methods consistently outperform traditional models in predicting educational outcomes. For example, [16] integrated Random Forest, K-Nearest Neighbor, and XGBoost using a stacking ensemble framework, achieving an R^2 score of 0.97 on benchmark educational datasets. Similarly,

[17] demonstrated that an AdaBoost-based ensemble model outperformed Decision Trees, Neural Networks, and SVMs in predicting academic success.

In addition, hybrid ensemble models combining boosting and bagging strategies have demonstrated remarkable results. [18] proposed a multi-class ensemble model that integrated Decision Tree, KNN, Naïve Bayes, and SVM classifiers, achieving 93% accuracy [19]. tested a supervised ensemble framework and confirmed that combining base classifiers improved classification performance across multiple student datasets [20]. introduced a graph-based ensemble learning system that improved accuracy by 14.8% compared to traditional single models.

Feature interpretability has become another critical dimension of recent studies [21]. designed a stacking ensemble with SVM, Random Forest, and AdaBoost and demonstrated that feature importance analysis enhances model transparency. Similarly, [22] applied Extreme Learning Machine models optimized via particle swarm algorithms, improving both performance and interpretability.

Moreover, ensemble learning has proven effective in diverse educational contexts, including online and multimedia-based learning [23]. combined CNN-derived features with Random Forest and SVM ensemble models to predict student performance in virtual environments with 98.99% accuracy. Likewise, [24] proposed an ensemble model integrating convolutional features and traditional ML algorithms, achieving 97.88% accuracy on learning management system data. Studies during the COVID-19 pandemic, such as those by. [25] and [26], also demonstrated that ensemble models could successfully predict student risk in distance learning systems.

The growing application of ensemble methods highlights their adaptability and effectiveness across educational settings. However, the literature still lacks large-scale applications using regional or national datasets, particularly in secondary education contexts. The present study contributes to filling this gap by implementing Random Forest and Gradient Boosting on a comprehensive state-level dataset, offering new insights into how ensemble-based machine learning can be scaled for educational policy and planning.

Methods

This study employed a supervised machine learning approach to predict student academic performance using comprehensive state-level secondary education data. The methodological framework followed a reproducible process consisting of data collection, preprocessing, exploratory data analysis, model development, evaluation, and interpretation which illustrated in figure 1. Two ensemble-based regression algorithms, Random Forest Regressor (RFR) and Gradient Boosting Regressor (GBR), were implemented due to their proven ability to model complex nonlinear relationships, handle large numbers of features, and provide interpretability through feature importance analysis. The main objective of this approach was to develop accurate predictive models while simultaneously identifying the key educational, demographic, and institutional factors influencing student achievement.

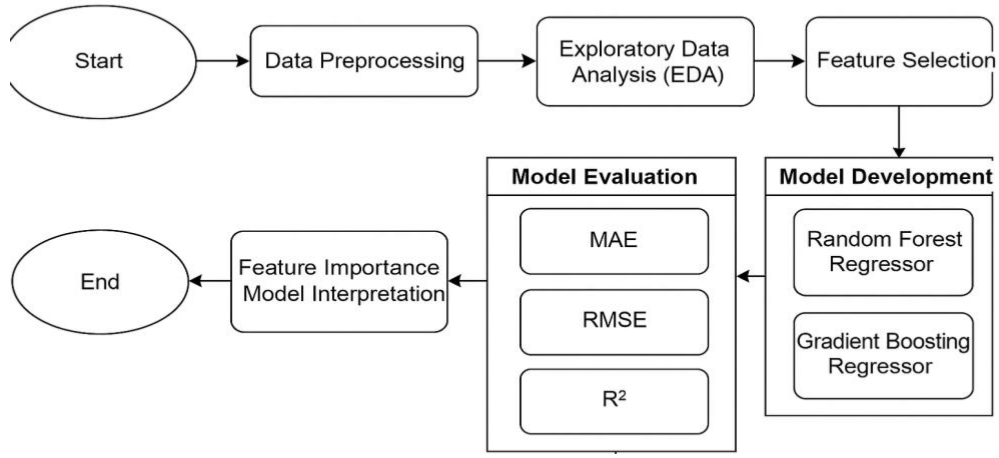

Figure 1 **Research Methods**

The dataset used in this research was obtained from the 2015–2016 Statewise Secondary Education Statistics, which provide comprehensive educational indicators for multiple states. Each record represented aggregated educational performance data, including examination participation rates, gender-based passing percentages, teacher distribution, infrastructure variables, and inclusion of students with special needs. The target variable used for prediction was the average student pass rate, which served as an indicator of overall academic performance at the state level. This dataset was selected for its diversity and representativeness of educational outcomes across regions.

Data preprocessing was performed to ensure data quality and analytical reliability. Initially, the dataset was examined for duplicates, inconsistencies, and missing values. Records with excessive missing entries were excluded, while remaining missing values in numerical columns were imputed using the median strategy through Scikit-learn's SimpleImputer, chosen for its robustness against outliers. Outliers were identified using the Interquartile Range (IQR) method, defined mathematically as:

$$Q_1 - 1.5(Q_3 - Q_1) \quad \boldsymbol{and} \quad Q_3 + 1.5(Q_3 - Q_1) \tag{1}$$

Values outside this range were capped or adjusted to minimize their influence on model performance. Since ensemble models are insensitive to scaling, normalization and standardization were not applied. Non-numeric features were label-encoded when necessary, and irrelevant administrative attributes such as state identifiers were removed. After cleaning, exploratory data analysis (EDA) was performed using Matplotlib and Seaborn to visualize feature distributions, identify potential correlations, and detect multicollinearity among variables.

Exploratory analysis revealed that gender-based participation and pass rates had strong positive correlations with overall academic performance, while institutional resources such as the number of teachers, librarians, and inclusive education programs were also associated with higher pass rates. To reduce redundancy, features exhibiting a Pearson correlation coefficient above 0.90 were removed. Feature selection was then refined using both statistical and model-based approaches. First, correlation analysis filtered redundant predictors. Subsequently, Random Forest's intrinsic feature importance attribute ranked the remaining features by their contribution to predictive accuracy. The most influential predictors included the percentage of general category girls passing class 10 (pass_g_gen_py10), the number of boys and girls appearing for exams (apr_b_gen_py10 and apr_g_gen_py10), the number of students with special needs in class 12 (cwsnc12_g), and the availability of librarians in schools (librarian_6).

Model development was conducted using two ensemble regression algorithms, Random Forest and Gradient Boosting, implemented via the Scikit-learn library in Python 3.12. The dataset was split into 80% for training and 20% for testing to evaluate generalization capability. The Random Forest Regressor is a bagging ensemble that constructs multiple independent decision trees on random subsets of the data and averages their outputs to minimize variance. The prediction for an input vector is defined as:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \tag{2}$$

$T_i(x)$ represents the prediction of the i^{th} decision tree, and N is the total number of trees. In contrast, the Gradient Boosting Regressor is a boosting algorithm that builds trees sequentially, where each new learner aims to correct the residuals of the previous model. The iterative update of the boosted model is expressed as:

$$F_m(x) = F_{m-1}(x) + \nu h_m(x) \tag{3}$$

$F_m(x)$ denotes the ensemble model after mmm iterations $h_m(x)$ represents the weak learner, and ν is the learning rate controlling the contribution of each tree. In this study, the Random Forest model used 100 estimators and a fixed random state of 42, while the Gradient Boosting Regressor used 100 estimators, a learning rate of 0.1, and a maximum tree depth of 3.

Model performance was evaluated using three regression metrics: MAE, RMSE and the coefficient of determination (R^2). These metrics were computed as follows:

$$\begin{gathered} \boldsymbol{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y_i}| \\ \boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y_i})^2} \\ R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{\sum_{i=1}^{n} \left(y_i - \underline{y}\right)^2} \end{gathered} \tag{4}$$

y_i is the observed value, $\hat{y_i}$ the predicted value, $\underline{y}$ the mean of observed outcomes, and n the number of samples. MAE measures the average magnitude of prediction errors, RMSE emphasizes larger deviations by penalizing significant errors, and R^2 quantifies the proportion of variance in the target variable explained by the model.

The results indicated that the Gradient Boosting model achieved an R^2 score of approximately 0.68, outperforming the Random Forest model, which obtained an R^2 score of about 0.58. These results suggest that Gradient Boosting better captured nonlinear dependencies among features, while Random Forest exhibited higher stability and interpretability. These outcomes are consistent with the literature, which reports that boosting algorithms generally outperform bagging models in complex educational datasets.

Finally, feature importance analysis was performed to interpret the contribution of each predictor to the model's performance. Random Forest's Gini importance was employed to quantify each feature's contribution, defined mathematically as:

$$I_G(f) = \sum_{t \in T} p(t) \Delta i(t, f) \tag{5}$$

$p(t)$ represents the proportion of samples reaching node t, and $\Delta\, i(t,\, f)$ is the impurity reduction after splitting on feature f. The resulting analysis highlighted pass_g_gen_py10, apr_b_gen_py10, apr_g_gen_py10, cwsnc12_g, and librarian_6 as the top predictors influencing student outcomes. Visualization of the top fifteen features (figure 3) demonstrated that gender participation, inclusivity, and institutional support were the most influential determinants of academic performance.

All analyses were implemented in a Jupyter Notebook environment using Python 3.12 on a system with an Intel Core i7 processor, 16 GB RAM, and a Windows 11 64-bit operating system. Libraries such as Pandas and NumPy were utilized for data manipulation, Matplotlib and Seaborn for visualization, and Scikit-learn for machine learning model implementation. The entire workflow ensured reproducibility, transparency, and scalability, making the approach adaptable for other large-scale educational datasets and applicable in data-driven education policy and planning.

Result and Discussion

The results of this study demonstrate that machine learning techniques can effectively predict student academic performance using large-scale state-level secondary education data. Two ensemble-based algorithms, Random Forest and Gradient Boosting, were trained and evaluated to determine their predictive strength and generalization capability. The evaluation utilized three standard regression metrics: MAE, RMSE and the coefficient of determination (R^2), which together provide a robust measure of model accuracy and consistency. As shown in figure 2, the Gradient Boosting model achieved an R^2 score of approximately 0.68, outperforming the Random Forest model, which obtained an R^2 value of about 0.58. This result indicates that Gradient Boosting successfully explained nearly 68 percent of the variance in student performance, while Random Forest accounted for approximately 58 percent. The higher explanatory power of Gradient Boosting suggests that it captures more complex nonlinear patterns and interactions among the predictor variables. Its superior performance likely stems from the algorithm's iterative optimization process, which incrementally minimizes prediction error by combining multiple weak learners into a stronger composite model.

Although the Random Forest model produced slightly lower predictive accuracy, it remains an important model in the analysis due to its robustness and interpretability. Random Forest's ensemble mechanism, which averages predictions from multiple decision trees, helps prevent overfitting and maintains stability across diverse feature sets. Its ability to handle high-dimensional educational data without extensive parameter tuning made it a reliable baseline for comparison. Moreover, Random Forest provided clear feature importance rankings that helped identify the factors contributing most to student performance. Together, both models demonstrated that ensemble learning methods are highly effective for educational data mining tasks, particularly when dealing with heterogeneous and multivariate datasets. The results also highlight that adopting advanced algorithms such as Gradient Boosting can substantially improve predictive precision, thereby offering greater potential for data-driven decision-making in educational policy and school management.

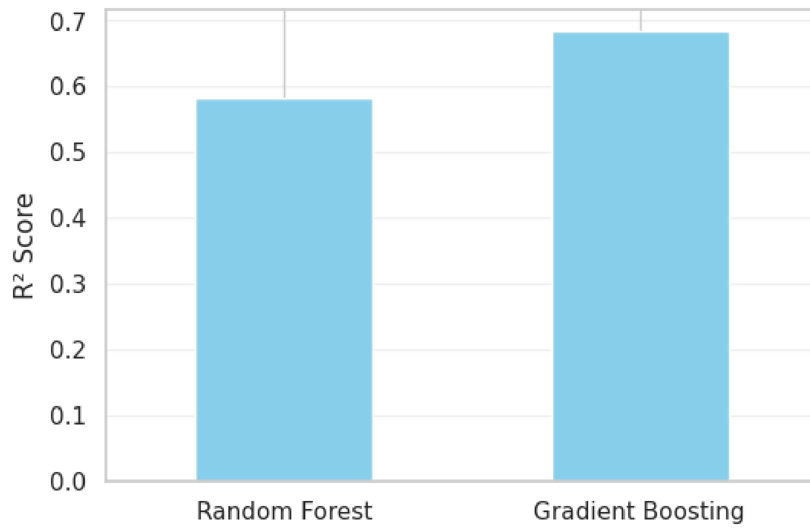

Figure 2 Model Comparison (R^2 Score) between Random Forest and Gradient Boosting

The analysis of feature importance derived from the Random Forest model provided deeper insight into the factors most strongly associated with student academic performance. As illustrated in figure 3, the model identified several features that contribute significantly to predictive accuracy, reflecting the multifaceted nature of educational outcomes. The most influential variables include pass_g_gen_py10 (percentage of general category girls passing class 10), apr_b_gen_py10 and apr_g_gen_py10 (numbers of male and female students appearing in exams), and cwsnc12_g (number of students with special needs enrolled in class 12). The prominence of these features suggests that student participation and success rates in examinations are powerful indicators of overall educational effectiveness. In particular, the inclusion of gender-segregated pass rates underscore the continued importance of gender as a determinant in academic achievement. The presence of cwsnc12_g as a strong predictor highlights the increasing role of inclusive education programs that support students with special needs, emphasizing that equitable access to schooling directly influences overall academic outcomes.

In addition to participation and inclusivity variables, several school infrastructure and staffing features were found to have substantial predictive weight. Among these, pass_sci_gen_b (science stream boys passing class 10), librarian_6 (availability of librarians in schools), and tchse_m (number of senior secondary male teachers) emerged as influential predictors. These results reveal that wellresourced schools with qualified staff and adequate academic facilities are more likely to achieve higher student performance levels. The strong contribution of the librarian and teacher-related variables reflects the importance of human capital and educational support systems in facilitating learning outcomes. Furthermore, the presence of science-related achievement indicators among the top features implies that subject-specific competencies, particularly in STEM areas, are critical for academic success in secondary education. Overall, the analysis demonstrates that student outcomes are shaped not only by individual participation but also by institutional support structures and educational resource allocation. These findings reinforce the value of using explainable machine learning approaches to uncover hidden patterns in educational data and to guide targeted interventions that can improve learning equity and quality across regions.

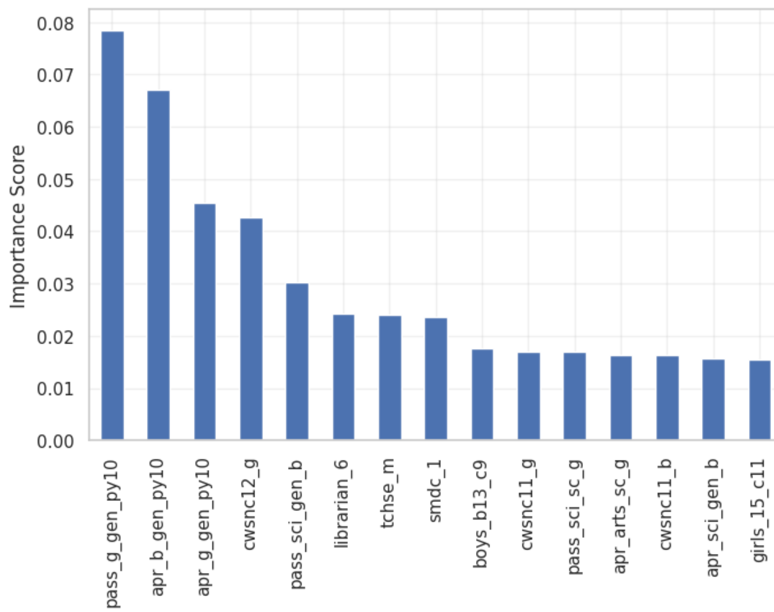

Figure 3 Top 15 Most Important Features in Predicting Student Performance (Random Forest)

In general, the results indicate that ensemble learning models, particularly Gradient Boosting, can effectively predict educational outcomes and reveal the key factors that influence performance. The findings suggest that improving student participation rates, enhancing teacher quality, and ensuring the availability of school resources such as libraries and laboratories can improve overall academic success. Moreover, the importance of inclusive education indicators, such as students with special needs, highlights the role of accessibility and equity in promoting better learning outcomes. These results provide valuable insights for policymakers and educational planners to design data-driven strategies that strengthen educational systems and reduce performance disparities between regions. Future research could enhance these models by incorporating socio-economic and demographic variables, such as household income, teacher training quality, and school funding, to achieve more comprehensive and generalizable predictions.

Conclusion

The findings of this study demonstrate that machine learning techniques, particularly ensemble-based models such as Random Forest and Gradient Boosting, can effectively predict student academic performance using large-scale secondary education data. The Gradient Boosting model achieved superior predictive accuracy with an R^2 score of approximately 0.68, indicating its strength in capturing nonlinear and complex relationships among educational indicators, while the Random Forest model provided valuable interpretability through its analysis of feature importance. The results revealed that factors related to exam participation, gender-based pass rates, teacher availability, school infrastructure, and inclusive education play critical roles in shaping student outcomes. Specifically, the prominence of female participation and success variables underscores the continuing benefits of gender-focused educational initiatives, and the importance of inclusive education variables highlights the value of equitable access for students with special needs. These insights suggest that policy interventions aimed at improving teacher quality, strengthening institutional resources, and promoting inclusive learning environments can lead to significant improvements in academic achievement. From a methodological perspective, this research confirms that artificial intelligence and data-driven analytics hold great potential for transforming educational policy and practice by providing evidence-based recommendations and predictive insights. Future work should integrate broader socio-economic and behavioral variables to enhance model robustness and apply explainable AI approaches to improve the interpretability and ethical use of predictive systems in education. Collectively, these findings contribute to the growing body of knowledge on the use of artificial intelligence in education and highlight the transformative potential of predictive modeling in achieving more equitable and effective learning outcomes.

Declarations

Author Contributions

Conceptualization: T.; Methodology: T.; Software: T.S.M.; Validation: D.F. and T.; Formal Analysis: T.S.M.; Investigation: T.S.M.; Resources: T.; Data Curation: T.; Writing Original Draft Preparation: T.S.M.; Writing Review and Editing: T. and T.S.M.; Visualization: T.S.M.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, andor publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Daware, P., “Regression and Ensemble Models for Analyzing Students Key Performance Factors,” *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 11, pp. 1437–1443, 2024, doi: 10.22214/ijraset.2024.65404.

[2] Sandra, L., Lumbangaol, F., dan Matsuo, T., “Machine Learning Algorithm to Predict Student’s Performance: A Systematic Literature Review,” *TEM Journal*, vol. 10, no. 4, pp. 1919–1927, 2021, doi: 10.18421/TEM104-56.

[3] Deepak Chowdary, E., Lakshmi Prasanna, V., Vamsi Krishna T., V., dan Yenduri, G., “Predicting Student Academic Performance with Ensemble Classification Method on Imbalanced Educational Data,” *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 4, pp. 1543–1551, 2019, doi: 10.35940/ijrte.D7741.118419.

[4] Wang, Y., Ding, A., Guan, K., Wu, S., dan Du, Y., “Graph-based Ensemble Machine Learning for Student Performance Prediction,” *arXiv preprint* arXiv:2112.07893, 2021, doi: 10.48550/arXiv.2112.07893.

[5] Balcıoğlu, Y. S., dan Artar, M., “Predicting academic performance of students with machine learning,” *Journal of Information Technology for Development*, vol. 41, no. 3, 2023, doi: 10.1177/02666669231213023.

[6] Abiodun, O. J., dan Wreford, A. I., “Student's Performance Evaluation Using Ensemble Machine Learning Algorithms,” *Engineering and Technology Journal*, vol. 9, no. 8, pp. 4768–4775, 2024, doi: 10.47191/etj/v9i08.23.

[7] Han, M., Tong, M., Chen, M., Liu, J., dan Liu, C., “Application of Ensemble Algorithm in Students' Performance Prediction,” dalam *Proceedings of the 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, Hamamatsu, Japan, vol. 2017, no. November, pp. 735–740, 2017, doi: 10.1109/IIAI-AAI.2017.73.

[8] Saluja, R., Rai, M., dan Saluja, R., “Designing new student performance prediction model using ensemble machine learning,” *Journal of Autonomous Intelligence*, vol. 6, no. 1, p. 583, 2023, doi: 10.32629/jai.v6i1.583.

[9] Emima, A., dan George Amalarethinam, D. I., “Integrative Ensemble Learning Algorithm for Predicting Students’ Performance,” *Indian Journal of Science and Technology*, vol. 18, no. 1, pp. 72–84, 2025, doi: 10.17485/IJST/v18i1.3718.

[10] Al-Ameri, A., Al-Shammari, W., Castiglione, A., Nappi, M., Pero, C., dan Umer, M., “Student Academic Success Prediction Using Learning Management Multimedia Data With Convoluted Features and Ensemble Model,” *ACM Journal of Data and Information Quality*, vol. 17, no. 3, Art. no. 12, pp. 1–16, 2025, doi: 10.1145/3687268.

[11] Tao, Y., Yang, Y., Qu, F., dan Liu, Y., “Student Performance Prediction Using Stacking Ensemble Learning,” dalam *Proceedings of the 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC)*, Wuhan, China, vol. 2024, no. March, pp. 1466–1469, 2024, doi: 10.1109/EIECC64539.2024.10929481.

[12] Santiketa, N., Chaikhan, S., Ninrutsirikun, U., dan Wattanakitrungroj, N., "Student Academic Performance Prediction using Machine Learning with Various Features and Scenarios," dalam *Proceedings of the 28th International Computer Science and Engineering Conference (ICSEC)*, Khon Kaen, Thailand, vol. 2024, no. December, pp. 1–6, 2024, doi: 10.1109/ICSEC62781.2024.10770729.

[13] Pawar, A., Shenoy, K. M., Prabhu, S., dan Rai, D. G., "Performance analysis of machine learning algorithms: Single Model VS Ensemble Model," *Journal of Physics: Conference Series*, vol. 2571, no. 1, Art. no. 012007, 2023, doi: 10.1088/1742-6596/2571/1/012007.

[14] Tong, T., dan Li, Z., "Predicting learning achievement using ensemble learning with result explanation," *PLOS ONE*, vol. 20, no. 1, Art. no. e0312124, 2025, doi: 10.1371/journal.pone.0312124.

[15] Yan, L., dan Liu, Y., "An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning," *Symmetry*, vol. 12, no. 5, p. 728, 2020, doi: 10.3390/sym12050728.

[16] Soares, W. L., dan Fagundes, R., "Extreme Learning Machine Models for Predicting Student Performance," *International Journal of Computer Applications*, vol. 174, no. 22, pp. 1–7, 2021, doi: 10.5120/ijca2021921122.

[17] Saidani, O., Umer, M., Alshardan, A., et al., "Student academic success prediction in multimedia-supported virtual learning system using ensemble learning approach," *Multimedia Tools and Applications*, vol. 83, no. March, pp. 87553–87578, 2024, doi: 10.1007/s11042-024-18669-z.

[18] M. Tsai-Lai and T. Hariguna, "Analyzing Company Hiring Patterns Using K-Means Clustering and Association Rule Mining: A Data-Driven Approach to Understanding Recruitment Trends in the Digital Economy," Journal of Digital Society, vol. 1, no. 1, Sep. 2025,

[19] Abdullah, M., Al-Ayyoub, M., Shatnawi, F., Rawashdeh, S., dan Abbott, R., "Predicting students' academic performance using e-learning logs," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 2, pp. 831–839, 2022, doi: 10.11591/ijai.v12.i2.pp831-839.

[20] Karalar, H., Kapucu, C., dan Gürüler, H., "Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system," *International Journal of Educational Technology in Higher Education*, vol. 18, no. December, Art., no. 63, 2021, doi: 10.1186/s41239-021-00300-y.

[21] Zhang, Z., dan Li, L., "Exploration of Intelligent Education: Student Performance Prediction," dalam *Proceedings of the Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Education Digitalization and Computer Science (EDCS)*, vol. 2024, no. September, pp. 201–210, 2024, doi: 10.1145/3686424.3686459.

[22] Shaik, S., Dasari, S. S., Singala, S., Mangalampalli, K., dan Varalakshmi, B., "ESN-MAXOUT: An Efficient Framework for Predicting Student's Academic Performance in Online Learning," *SN Computer Science*, vol. 7, no. 1, p. 55, 2025, doi: 10.1007/s42979-025-04618-6.

[23] Kaviyarasi, R., dan Balasubramanian, T., "Predictive analysis of academic performance of college students using ensemble stacking," *Kongunadu Research Journal*, vol. 7, no. 2, pp. 94–98, 2020, doi: 10.26524/krj.2020.28.

[24] Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., dan Durán-Domínguez, A.,

"Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review," *Applied Sciences*, vol. 10, no. 3, p. 1042, 2020, doi: 10.3390/app10031042.

[25] V. K. P, P. S, and M. Batumalay, "Examining the Association Between Social Media Use and Self-Reported Social Energy Depletion: A Machine Learning Approach," Journal of Digital Society, vol. 1, no. 3, Sep. 2025, doi: 10.63913/jds.v1i3.50.

[26] Borno, Z. S., Sakib, N., dan Anwar, S. S., "Performance Analysis of Ensemble Machine Learning Algorithms in PDF Malware Detection," dalam *Proceedings of the IEEE 9th International Women in Engineering Conference on Electrical and Computer Engineering (WIECON-ECE)*, Thiruvananthapuram, India, vol. 2023, no. March, pp. 195–200, 2023, doi: 10.1109/WIECON-ECE60392.2023.10456385.