Comparative Analysis of Multiple Linear Regression and Random Forest Regression in Predicting Academic Performance of Students in Higher Education

.

**ABSTRACT**

|  |
| --- |
| **Aims:** This study aimed to compare the predictive accuracy of Multiple Linear Regression (MLR) and Random Forest Regression (RFR) models in forecasting academic performance among Social Work students. Specifically, it sought to identify which among the considered variables—study habits, learning styles, stress, anxiety, coping mechanisms, study motivation, exam preparation, age, and parental income—most significantly influenced students’ academic outcomes.  **Study Design:**  Quantitative predictive-correlational research methodology.  **Place and Duration of Study:**  The research was conducted with 45 Social Work students enrolled in a Statistics course at a higher education institution in the Philippines during the second semester of the 2024–2025 academic year.  **Methodology:** Validated questionnaires were distributed via Google Forms to assess students’ behaviors and emotional states. Instrument reliability was confirmed using JAMOVI (version 2.6.23). Descriptive statistics were applied to evaluate mean scores, while Spearman’s rho was used for correlation analysis due to violations of normality. Predictive modeling was conducted using Python (Jupyter Notebook), employing both MLR and RFR to evaluate predictive power and identify significant factors. Model performance was assessed using the Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R²).  **Results:** Descriptive analysis showed moderate levels of stress, anxiety, and coping mechanisms, with exam preparation yielding the lowest mean. Anxiety level displayed a moderate negative correlation with academic performance. Both MLR and RFR models identified Learning Style, Exam Preparation, and General Stress as the strongest predictors. The RFR model outperformed MLR, achieving a lower MAPE (2.9481) and higher R² (0.865) compared to MLR (MAPE = 3.3690, R² = 0.216), indicating stronger predictive accuracy and better handling of nonlinear relationships.  **Conclusion:** The Random Forest model outperformed the Multiple Linear Regression model in predicting academic performance, demonstrating higher accuracy and a stronger capacity to model nonlinear relationships. Learning Style, Exam Preparation, and General Stress emerged as the most influential predictors. These findings support the integration of machine learning approaches in educational research and suggest that tailored learning strategies, structured exam preparation, and stress management interventions may contribute to improved academic outcomes. |

*Keywords: Academic Performance, Higher Education, Multiple regression, Random Forest Regression, Machine Learning, Study Habits, Academic Stress, Resilience, Teaching Style, Mean Absolute Percentage*

**1. INTRODUCTION**

Globally, academic achievement serves as a crucial metric of educational accomplishment, influencing both individual prospects and the overall standing and efficacy of higher education institutions. It functions as an indicator of a student's proficiency in mastering curriculum, applying knowledge, and excelling in demanding academic settings. Additionally, it offers educational stakeholders—such as administrators, professors, and policymakers—criteria to assess instructional quality, pinpoint curricular deficiencies, and evaluate institutional efficacy. In a progressively globalized and competitive educational environment, academic success is crucial in securing scholarships, graduate programs, and employment possibilities. Although academic achievement has always been associated with intellectual aptitude and cognitive abilities, there is an increasing agreement that this viewpoint is insufficient. Recent research indicate that academic outcomes are influenced by non-cognitive elements, including emotional regulation, motivation, learning environments, and the quality of student-instructor relationships (Espericueta-Medina et al., 2020; Dallasheh, 2024; Zayed, 2024). These findings are instigating a paradigm change from only quantitative assessments of intelligence to more comprehensive, student-centered methodologies for educational evaluation and intervention (Rafiq et al., 2025; Zheng et al., 2024).

In the Philippine context, this multifaceted dilemma presents itself distinctly, especially in private tertiary schools located in rural or provincial regions. These institutions, frequently functioning with restricted financial resources and facilities, cater to students encountering diverse socio-economic challenges. Financial instability, limited internet access, domestic responsibilities, and mental health challenges are prevalent, particularly in places with inadequate or nonexistent educational support systems. Students enrolled in emotionally taxing programs, such as Social Work, encounter heightened stress as a result of their direct involvement with underprivileged communities through fieldwork and community-based education. These academic experiences, albeit pedagogically enriching, immerse students in emotionally intense situations that may reflect their personal life conditions (Zayed, 2024; RaWIv et al., 2024). The simultaneous challenges of academic and personal stress provide a multifaceted educational experience that necessitates thorough examination of the intricate aspects influencing academic success in these contexts. It is essential to identify which variables—such as stress management, resilience, study habits, and instructor empathy—most significantly impact student success.

Notwithstanding this complexity, numerous local and institutional studies persist in employing conventional statistical methods, such as Multiple Linear Regression (MLR), to analyze academic achievement. Although these methods possess some advantages, they are constrained in their capacity to identify nonlinear correlations and interaction effects among predictors (Gomes et al., 2020; Mbunge et al., 2021). This is a challenge for analyzing complex educational processes where factors do not function independently. The influence of stress on academic performance may be alleviated by resilience or intensified by an unsupportive learning environment—dynamics that conventional models may not adequately identify. As a result, there is growing interest in utilizing machine learning (ML) models such as Random Forest Regression, which are engineered to analyze extensive datasets, identify intricate patterns, and prioritize variables based on their predictive significance (Casinillo & Laurente, 2024; Amraouy et al., 2025). These methodologies have demonstrated efficacy in global educational contexts but are still inadequately employed in Philippine research, especially in Social Work education at provincial schools.

This is particularly significant when national educational bodies, such as the Commission on Higher Education (CHED), underscore the importance of innovation and evidence-based practices in facilitating educational transformation. CHED's policies promote research-based practices that improve learning outcomes, increase instructional delivery, and diminish equity inequalities in access and achievement. Their emphasis on regional and marginalized institutions highlights the necessity for localized, data-driven research to influence policy and practice. This study aimed to compare the predictive accuracy of Multiple Linear Regression and Random Forest Regression models in forecasting academic performance among Social Work students. Specifically, it sought to identify which among the considered variables—study habits, learning styles, stress, anxiety, coping mechanisms, study motivation, and exam preparation—were most influential in predicting students’ academic outcomes. The results of this comparison will help educators identify key areas for academic support and will demonstrate the effectiveness of machine learning models like Random Forest in educational research.

This study has three key objectives. The primary objective is to determine the levels of various psychological, behavioral, and academic factors that may influence student academic performance, including general stress, anxiety level, coping mechanism, study habits, learning style, study motivation, and exam preparation. The secondary objective is to predict students’ academic performance (grades) based on these identified factors using two modeling approaches: Multiple Linear Regression and Random Forest Regression. Finally, the last objective is to compare the predictive accuracy of the two models using key evaluation metrics, with a particular focus on the Mean Absolute Percentage Error (MAPE) as the basis for determining the more effective model. This study's results aim to furnish educators, administrators, and policymakers with practical insights for enhancing academic support systems, optimizing pedagogical strategies, and developing curricula that are more responsive to the distinct needs of students in provincial Social Work programs.

**2. methodology**

**2.1 Research Design**

This study employed a quantitative, predictive-comparative research approach to investigate the factors influencing academic achievement among higher education students. It assessed the predictive accuracy of Multiple Linear Regression (MLR) and Random Forest Regression (RFR) in identifying key variables, including academic stress, study habits, and resilience. Data were collected using a standardized questionnaire. Pilot testing was conducted using JAMOVI, while the final data analysis and model comparison were carried out in Python. Model performance was evaluated using Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R²) to assess both accuracy and explanatory power.

**2.2 Research Respondents**

The research encompassed 45 social work students in higher education from a singular cohort at a private university in Misamis Occidental. Participants were recruited through purposive sampling based on their relevance to the study's aim of finding predictors of academic performance. Ethical principles, encompassing secrecy and voluntary involvement, were rigorously maintained.

**2.3 Instrument of the Study**

This study utilized the **Study Habits and Learning Styles Questionnaire** and the **Stress and Anxiety Assessment Questionnaire**. The surveys underwent a reliability assessment using **Cronbach’s Alpha** in **JAMOVI software (version 2.6.23).** All subscales produced alpha values ranging from **0.70 to 1.00,** signifying that the instruments employed in this study exhibit **acceptable to exceptional internal consistency**. Consequently, the instruments are considered dependable for evaluating the intended constructs.

**2.3.1 Study Habits and Learning Styles Questionnaire**

This tool evaluates students' study practices, learning preferences, academic motivation, and test preparation tactics through four components: Study Habits, Learning Styles, Study Motivation, and test Preparation. Each component comprises 10 items evaluated on a 4-point Likert scale (1 = Strongly Disagree to 4 = Strongly Agree). The outcomes are analyzed utilizing established mean ranges to classify responses from Very Poor to Excellent.

**Table 1. Verbal Interpretation for Study Habits, Learning Styles, Study Motivation, and Exam Preparation**

|  |  |  |
| --- | --- | --- |
| Scale | Mean Range | Descriptive Rating |
| 4 | 3.25 – 4.00 | Excellent |
| 3 | 2.50 – 3.24 | Good |
| 2 | 1.75 – 2.49 | Poor |
| 1 | 1.00 – 1.74 | Very Poor |

**2.3.2 Stress and Anxiety Assessment Questionnaire**

This tool evaluates students' reported stress, anxiety, and coping techniques through three components: General Stress Levels, Anxiety Levels, and Coping Mechanisms. Each component comprises 10 items evaluated on a 4-point Likert scale (1 = Strongly Disagree to 4 = Strongly Agree), with results analyzed through mean ranges to categorize responses from Very Low to High or Effective.

**Table 2. Verbal Interpretation for Stress, Anxiety, and Coping Mechanisms**

|  |  |  |
| --- | --- | --- |
| Scale | Mean Range | Descriptive Rating |
| 4 | 3.25 – 4.00 | High / Effective |
| 3 | 2.50 – 3.24 | Moderate |
| 2 | 1.75 – 2.49 | Low / Poor |
| 1 | 1.00 – 1.74 | Very Low / Very Poor |

**2.3.3 Academic Performance**

Academic performance was assessed using student’s final average grades from the preceding semester. The grades were acquired with consent and utilized as the criterion variable to determine whether factor—study habits, learning styles, or stress/anxiety—serves as the most significant predictor of academic achievement.

**2.4 Data gathering procedure**

Data were gathered from 45 students utilizing two validated instruments: the Study Habits and Learning Styles Questionnaire and the Stress and Anxiety Assessment Questionnaire. Both instruments were transformed into Google Forms to guarantee accessibility and convenience. The link was disseminated to the participants, who were allotted sufficient time to complete the forms autonomously. The researcher observed the responses, alerted students to any incomplete items, and confirmed that all submissions were finalized prior to commencing data analysis.

**2.5 Statistical Treatment of Data**

This study's data was analyzed employing descriptive statistics, inferential statistics, and machine learning methodologies. Descriptive statistics, namely the mean, were employed to quantify general stress, anxiety, coping strategies, study habits, learning styles, study motivation, and exam preparation, which were analyzed using established linguistic categories. Multiple linear regression was utilized for inferential analysis to ascertain the degree to which these variables predicted academic performance and to find the components that significantly influenced students' academic achievements. To augment the reliability of the prediction, a Random Forest model an ensemble machine learning algorithm was employed. This method facilitated a more adaptable and precise forecast of academic achievement while pinpointing the most significant elements via feature relevance. The integration of traditional statistics with machine learning yielded a thorough study of the data.

**2.6 Model Evaluation**

This section describes the metrics employed to evaluate the efficacy of the predictive models. The Mean Absolute Percentage Error (MAPE) assessed predicting accuracy, whilst the coefficient of determination (R²) analyzed the models' effectiveness in explaining variance in academic achievement.

2.6.1 Mean Absolute Percentage Error (MAPE) as a Forecasting Accuracy Metric

This table categorizes the forecasting accuracy based on the Mean Absolute Percentage Error (MAPE), a widely used metric in time series forecasting (Liu & Zhang, 2024; Yang & Li, 2024). Lower MAPE values indicate higher accuracy, with thresholds defining different levels of forecasting performance.

Table 3. Forecasting Accuracy Classification Based on MAPE

|  |  |
| --- | --- |
| **MAPE** | **Forecasting Criterion** |
| 4.21-5.00 | Forecasting accuracy is very good |
| 3.41-4.20 | Forecasting accuracy is good |
| 2.61-3.40 | Forecasting accuracy is good enough |
| 1.81-2.60 | Forecasting accuracy is not good |
|  |  |
|  |  |

2.6.2 Coefficient of Determination (R²) As A Model Fit Metric

The coefficient of determination (R²) is a conventional metric employed to assess the explanatory capacity of prediction models. It measures the extent to which the variance in the dependent variable may be anticipated from the independent factors. An R² value spans from 0 to 1, with values approaching 1 signifying a more robust model fit. This study employed R² in conjunction with MAPE to evaluate the accuracy and explanatory power of both the Multiple Linear Regression (MLR) and Random Forest Regression (RFR) models.

The interpretation of R² values for model fit is informed by the thresholds typically utilized in educational and behavioral sciences, as suggested by Chicco et al. (2021) and endorsed by Hair et al. (2014) for structural model assessment:

Table 4: Model Fit Classification Based on R²

|  |  |
| --- | --- |
| Value Range | Interpretation |
|  | Strong model fit |
| 0.50 – 0.74 | Moderate model fit |
|  | Weak model fit |

3. RESULTS AND DISCUSSION

This section reveals the results obtained via descriptive statistics, Multiple Linear Regression (MLR), and the Random Forest Regression (RFR) model. The descriptive analysis offers a summary of students' stress levels, anxiety, coping strategies, study habits, learning preferences, motivation, and exam preparation techniques. Both regression models were utilized to analyze the correlations between these variables and academic performance. The MLR model found critical variables significantly linked to academic success, whereas the RFR model enhanced predicted accuracy by capturing intricate, nonlinear patterns in the data. All analyses were performed via Jupyter Notebook, and model efficacy was assessed through the Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R²), where reduced MAPE and elevated R² values signify superior model fit. These integrated analytical methodologies facilitated the identification of the primary factors of student academic progress.

3.1 Levels of psychological, behavioral, and academic factors affecting student performance.

Table 5 revealed the mean scores, standard deviations, and verbal interpretations of the primary variables considered as predictors of academic performance among 45 Social Work students enrolled in a Statistics course. These include general stress, anxiety levels, coping mechanisms, study habits, learning styles, study motivation, and exam preparation. Understanding student performance across these domains provides valuable insight into the behavioral and psychological factors that may influence academic achievement in a cognitively demanding subject like Statistics.

Coping Mechanism recorded the highest mean score of 3.24 (SD = 0.48), suggesting that most students moderately apply effective strategies to manage stress and academic challenges. This implies a generally adaptive approach to academic pressures, allowing students to maintain focus and emotional regulation. According to Pascoe et al. (2020) and Huerta (2022), effective coping mechanisms act as buffers against academic stress and help students remain productive in high-pressure learning environments. Furthermore, Musso et al. (2020) emphasized that students with better emotional self-regulation tend to demonstrate greater persistence and perform better academically, particularly in mentally intensive subjects.

In contrast, Exam Preparation had the lowest mean score of 2.60 (SD = 0.34), though still categorized as “good.” This relatively low score suggests that while students may manage stress adequately, they may be underprepared for assessments due to ineffective strategies or time management. Pekrun and Stephens (2012) highlighted that structured and consistent exam preparation significantly enhances academic outcomes. Similarly, von der Embse et al. (2018) linked insufficient preparation and test-related anxiety to underperformance in quantitative courses, including mathematics and statistics. The contrast between high coping and low exam preparation highlights a disconnect between emotional resilience and actual academic behavior. While students may be capable of managing pressure, they may not be translating that into effective study actions.

This reflects the findings of Radišić et al. (2022), who argued that integrating strong study habits with effective learning approaches is crucial for bridging performance gaps. Likewise, Popoola and Hendricks (2014) noted that aligning learning styles with instructional methods, alongside the presence of academic support systems, can enhance student engagement and academic achievement. In light of these findings, improvements in exam readiness, personalized learning strategies, and academic support programs may be beneficial for students navigating challenging subjects. As Pascarella and Terenzini (2005) suggested, initiatives that reduce anxiety and boost motivation can lead to measurable gains in student performance.

Table 5. Descriptive Statistics and Verbal Interpretation of Academic Predictors Among Social Work Students

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean | Standard Deviation | Verbal Interpretation |
| General Stress | 2.96 | 0.48 | Moderate Stress |
| Anxiety Level | 2.82 | 0.43 | Moderate Anxiety |
| Coping Mechanism | 3.24 | 0.48 | Moderate Coping Strategies |
| Study Habits | 2.71 | 0.30 | Good Study Habits |
| Learning Style | 2.85 | 0.27 | Moderate Preference for Active Learning Styles |
| Study Motivation | 2.89 | 0.30 | Moderate Motivation |
| Exam Preparation | 2.6 | 0.34 | Good Exam Preparation |
|  |  |  |  |

**3.2 Comparison of Predictive Models and Identification of Key Academic Performance Factors**

Prior to evaluating the predictive efficacy of the models, a Spearman's rho correlation study was performed to determine the degree and direction of the association between each academic element and student performance, as shown by grades. The non-parametric method was chosen based on the Shapiro-Wilk test results, which indicated that several variables, including Grades, General Stress, and Learning Style, were not regularly distributed.  
  
The correlation findings are encapsulated in Table 6. Anxiety Level demonstrated the most significant link among all predictors, revealing a moderate negative correlation with grades (ρ = −0.331, p < 0.05). This indicates that individuals with elevated anxiety levels were more prone to diminished academic performance. Additional variables, including Learning Style (ρ = −0.163), General Stress (ρ = −0.060), and Study Motivation (ρ = −0.037), exhibited weak or negligible relationships with academic performance. Variables such as Study Habits and Coping Mechanism exhibited negligible positive relationships that lacked statistical significance.

**Table 6 Spearman Correlation Between Academic Factors and Student Grades**

|  |  |  |
| --- | --- | --- |
| Academic Performance Factor | Spearman rho value | p-value |
| Coping Mechanism | 0.051 | 0.743 |
| Study Habits | 0.047 | 0.762 |
| Study Motivation | -0.037 | 0.81 |
| General Stress | -0.06 | 0.698 |
| Learning Style | -0.163 | 0.291 |
| Anxiety Level | -0.331 | 0.028 |

After this preliminary investigation, Multiple Linear Regression (MLR) and Random Forest Regression (RFR) models were constructed utilizing the identical collection of covariates to evaluate their overall efficacy in predicting academic performance. The models were assessed using two primary metrics: Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R²).

Table 7 presents the comparative results of two predictive models—Random Forest Regression (incorporating feature importance) and Multiple Linear Regression—used to assess the influence of various psychological, behavioral, and demographic factors on academic achievement.

The Random Forest model achieved a Mean Absolute Percentage Error (MAPE) of 2.9481 and a R² value of 0.865, demonstrating a robust ability to elucidate the variance in student grades when assessed on the entire dataset. In contrast, the Multiple Linear Regression model produced a MAPE of 3.3690 and a diminished R² value of 0.216, indicating a restricted explanatory capacity. These findings corroborate the research of Nachouki et al. (2023) and Batool et al. (2023), who highlighted the enhanced predictive efficacy of Random Forest in educational settings, notably owing to its resilience in modeling intricate, non-linear interactions among variables.

Both models consistently identified Learning Style and Exam Preparation as the primary predictors of academic performance. In the Random Forest model, Learning Style exhibited the highest feature relevance score (0.3455), succeeded by Exam Preparation (0.1647) and General Stress (0.1452). In the Linear Regression model, Learning Style exhibited the highest standardized coefficient (0.1904), closely succeeded by Exam Preparation (0.1875), underscoring their significant influence on academic achievements. These findings resonate with those of Falát and Piscová (2022) and Namoun and Alshanqiti (2020), who similarly underscored the significance of organized learning tactics and examination preparedness.  
  
Although there was a common focus on primary predictors, the models differed in their assessment of additional factors. For example, Study Motivation demonstrated a moderate coefficient (0.1041) in the Linear Regression model, suggesting a possible linear correlation with academic success. Nonetheless, its relatively lower feature relevance score (0.0494) in the Random Forest model indicates that its impact lessens when considering non-linear interactions. This comparison corroborates the findings of Khan and Ghosh (2021), who likewise noted variances in variable significance between traditional and machine learning models.  
  
Conversely, Coping Mechanism was the least significant variable in both models, exhibiting a feature importance of 0.0336 in the Random Forest model and a coefficient of 0.0157 in the Linear Regression model. This indicates limited significance in anticipating academic outcomes, aligning with the findings of Musso et al. (2020), who documented analogous results in studies focused on student-centered performance predictions.

The Random Forest model surpassed the traditional regression model, emphasizing Learning Style, Exam Preparation, and General Stress as the primary factors influencing academic success. The findings are corroborated by an expanding corpus of work about the utilization of machine learning in education, including Liu and Zhang (2024), who validated the enhanced predictive capability and interpretability of Random Forest models compared to conventional methods in educational data analytics.

Table 7. Comparison of Random Forest and Multiple Linear Regression in Predicting Academic Performance

|  |  |  |
| --- | --- | --- |
| Factor | Feature Importance  (Randon Forest) | Linear Regression |
|  | 0.865 | 0.216 |
| MAPE | 2.9481 | 3.3690 |
| Learning Style | 0.3455 | 0.1904 |
| Exam Preparation | 0.1647 | 0.1875 |
| General Stress | 0.1452 | 0.1442 |
| Income of Parents (Monthly) | 0.0828 | 0.0956 |
| Age | 0.0727 | 0.1015 |
| Anxiety Level | 0.0646 | 0.1035 |
| Study Motivation | 0.0494 | 0.1041 |
| Study Habits | 0.0415 | 0.0575 |
| Coping Mechanism | 0.0336 | 0.0157 |

**3.3Visualization of Model Predictions and Error Distribution**

This section illustrates the visual comparison of the prediction results from the Random Forest and Multiple Linear Regression models. The graphs depict the degree of alignment between each model's projected values and the actual student grades, facilitating a more precise evaluation of each model's accuracy and error distribution.

The prediction plots further demonstrate the advantages and drawbacks of the two employed models. The figure 1 indicates that forecasts frequently overestimate high-achieving students while underestimating low-achieving students, a trend commonly observed in the literature. This systemic bias arises from the model's failure to encapsulate intricate and nonlinear interactions, which are often evident in student performance data (Yesugade et al., 2024; Frick et al., 2023; Tzenios, 2020). Research indicates that linear models are inadequate in addressing latent variables such as motivation or engagement (Hilbert et al., 2021) and falter in managing variable interactions (Li, 2020; Cheng, 2024), rendering them less dependable across varied student populations.

Conversely, figure 2 exhibits a more accurate correlation between projected and actual grades. Data points are more closely grouped along the optimal prediction line, suggesting that prediction errors are uniformly distributed across all performance levels. This result underscores the model's capacity to handle nonlinear patterns and intricate variable interactions, enhancing its accuracy. The ensemble characteristic of Random Forest mitigates overfitting while maintaining generalizability. Research corroborates this benefit, with investigations by Frick et al. (2023), Jin (2023), and Rico-Juan et al. (2024) validating Random Forest’s exceptional efficacy in educational data modeling. These findings substantiate the application of sophisticated machine learning methodologies, such as Random Forest, in educational research, especially for generating precise, equitable, and nuanced forecasts of academic results.

Figure 1. Actual vs. Predicted Grades using Linear Regression

A graph with a line and dots

AI-generated content may be incorrect.

Figure 2. Actual vs. Predicted Grades using Random Forest

A graph with a line and blue dots

AI-generated content may be incorrect.

4. Conclusion

This study aimed to evaluate the predictive accuracy of Multiple Linear Regression (MLR) and Random Forest Regression (RFR) models in forecasting academic achievement among Social Work students. The objective was to ascertain which of the analyzed variables—study habits, learning styles, stress, anxiety, coping mechanisms, study motivation, exam preparation, age, and parental income—most significantly impacted academic performance. Data were gathered from 45 students enrolled in a Statistics course utilizing verified survey instruments and analyzed employing both conventional statistical approaches and machine learning techniques.

Descriptive data revealed that students shown moderate levels of stress, anxiety, and coping mechanisms, although exam preparation produced the lowest mean, underscoring a probable deficiency in academic preparedness. link research indicated that Anxiety Level exhibited a moderate negative link with academic performance, whilst other factors demonstrated weak or nonexistent associations.

Both prediction models highlighted Learning Style, Exam Preparation, and General Stress as the primary determinants of academic success. The Random Forest model demonstrated a lower Mean Absolute Percentage Error (MAPE = 2.9481) and a higher R² value (0.865) than the Multiple Linear Regression model (MAPE = 3.3690, R² = 0.216), signifying enhanced predictive accuracy and a greater capacity to elucidate nonlinear relationships within the data. These findings align with the literature highlighting the effectiveness of ensemble machine learning techniques in educational data processing.

Given these findings, it is advisable for academic institutions to emphasize tailored learning strategies, organized exam preparation assistance, and stress alleviation programs. Pragmatic interventions, including workshops on effective study methodologies, learner-centered pedagogical practices, and stress management initiatives, may enhance academic achievement. Future study ought to increase the sample size, investigate supplementary psychological and institutional variables, and utilize a wider array of machine learning models to improve prediction accuracy and generalizability across various educational environments.

Consent

Permission to conduct the study was sought from the appropriate institutional authorities. The researcher ensured that participation in the study was entirely voluntary, and that informed consent was obtained from all student participants prior to data collection. Participants were assured that their responses would remain confidential and anonymous, with no names or personal identifiers included in any part of the analysis or reporting. The purpose, procedures, and scope of the study were clearly explained to all participants. They were informed of their right to refuse participation or withdraw from the study at any point without any academic or personal consequences. Data was collected only after securing consent, and the researcher-maintained transparency and honesty throughout the research process. Ethical guidelines for conducting research involving human participants were strictly followed.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**References**

Acar, F. (2023). Underlying reasons behind the achievement of successful university students: A phenomenological inquiry. Journal of Pedagogical Sociology and Psychology, 5(2), 52–64. <https://doi.org/10.33902/JPSP.2023213916>

Adesokan, A. (2022). Emotional intelligence and academic performance of students with learning disability [Preprint]. ResearchGate. <https://www.researchgate.net/publication/369272923_Emotional_Intelligence_and_Academic_Performance_of_Students_with_Learning_Disability>

Anwar, A., Rehman, I. U., & Nasralla, M. M. (2023). Sentiment analysis and student emotions: Improving satisfaction in online learning platforms. In Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT), Orem, Utah, USA, July 10–13, 2023 (pp. 220–224). IEEE. <https://doi.org/10.1109/ICALT58600.2023.10293422>

Batool, S., Rashid, J., Nisar, M. W., Kim, J., & Kwon, H. Y. (2023). *Educational data mining to predict students' academic performance: A survey study*. Education and Information Technologies. <https://link.springer.com/article/10.1007/s10639-022-11152-y>

Cheng, Y. (2024). *Research on higher vocational students' reluctance to advance using Random Forest and SEM*. Advances in Engineering Innovation. <https://www.ewadirect.com/journals/aei/article/view/13245>

Falát, L., & Piscová, T. (2022). *Predicting GPA of university students with supervised regression machine learning models*. Applied Sciences, 12(17), 8403. <https://www.mdpi.com/2076-3417/12/17/8403>

Frick, S., Schmid, L., Kuhn, J. T., & Doebler, P. (2023). *Using multivariate random forests for predicting learning trajectories from digital training data*. OSF Preprints. <https://osf.io/x9vbh>

Hilbert, S., Coors, S., Kraus, E., Bischl, B., & Lindl, A. (2021). *Machine learning for the educational sciences*. Review of Education, 9(4), 803–828. <https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1002/rev3.3310>

Huerta, A. H. (2022). *Exploring undergraduate students' emotional vulnerability in Men of Color programs*. Journal of College Student Development, 63(2), 150–165. <https://muse.jhu.edu/article/853531>

Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). *A meta-analysis of the effects of audience response systems on cognition and affect*. Computers & Education, 94, 102–119. <https://www.sciencedirect.com/science/article/pii/S0360131515300853>

Jin, X. (2023). *Predicting academic success: A machine learning analysis of student, parental, and school efforts*. Asia Pacific Education Review. <https://link.springer.com/article/10.1007/s12564-023-09915-4>

Khan, A., & Ghosh, S. K. (2021). *Student performance analysis and prediction in classroom learning: A review of educational data mining studies*. Education and Information Technologies, 26, 6719–6758. <https://link.springer.com/article/10.1007/s10639-020-10230-3>

Kumar, S., Agarwal, M., & Agarwal, N. (2021). *Defining and measuring academic performance of HEI students: A critical review*. International Journal of Research in Education and Mathematics. <https://www.researchgate.net/publication/359002367_Defining_And_Measuring_Academic_Performance_of_HEI_Students>

Li, L. (2020). Causal effect random forest of interaction trees for learning individualized education strategies [Doctoral dissertation, University of Pittsburgh]. ProQuest Dissertations Publishing. <https://www.proquest.com/openview/6a8defced0c948dcb553c68a593860ba>

Liu, A., & Zhang, Y. (2024). *An efficient spatial-temporal transformer with temporal aggregation and spatial memory for traffic forecasting*. Expert Systems with Applications, 239, 122873. <https://doi.org/10.1016/j.eswa.2024.122873>

Liu, Y., & Zhang, Z. (2024). *Minute-level ultra-short-term power load forecasting based on time series data features*. Applied Energy, 362, 121184. <https://doi.org/10.1016/j.apenergy.2024.121184>

Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). *Predicting key educational outcomes in academic trajectories: A machine-learning approach*. Higher Education, 79(1), 1–22. <https://link.springer.com/article/10.1007/s10734-020-00520-7>

Nachouki, M., Mohamed, E. A., & Mehdi, R. (2023). *Student course grade prediction using the random forest algorithm: Analysis of predictors' importance*. Education and Information Technologies. <https://www.sciencedirect.com/science/article/pii/S2211949323000170>

Namoun, A., & Alshanqiti, A. (2020). *Predicting student performance using data mining and learning analytics techniques: A systematic literature review*. Applied Sciences, 11(1), 237. <https://www.mdpi.com/2076-3417/11/1/237>

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2020). *The impact of stress on students in secondary school and higher education*. International Journal of Adolescence and Youth, 25(1), 104–112. <https://doi.org/10.1080/02673843.2019.1596823>

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. San Francisco, CA: Jossey-Bass.

Pekrun, R., & Stephens, E. J. (2012). *Academic emotions and student engagement: Predicting success in school*. Educational Psychology Review, 24(2), 357–386. <https://doi.org/10.1007/s10648-012-9193-8>

Popoola, S., & Hendricks, C. S. (2014). *Learning styles of first-semester baccalaureate nursing students: A literature review*. Institute for Learning Styles Journal, 1, 1–9. <http://www.auburn.edu/academic/education/ilsrj/Journal%20Volumes/Fall%202014%20Vol%201%20PDFs/Learning%20Styles%20of%20First%20Semester%20Nursing%20Students%20Popoola%20and%20Hendricks.pdf>

Radišić, J., Hansen, K. Y., Ding, Y., & Liu, X. (2022). *Contextual effects on students’ achievement and academic self-concept in the Nordic and Chinese educational systems*. Large-scale Assessments in Education, 10(1), 1–20. <https://doi.org/10.1186/s40536-022-00133-9>

Rico-Juan, J. R., Cachero, C., & Macià, H. (2024). *Learning performance prediction using personality and LMS data*. Applied Intelligence. <https://link.springer.com/article/10.1007/s10489-024-05483-1>

Tzenios, N. (2020). *Examining the impact of EdTech integration on academic performance using Random Forest Regression*. ResearchBerg Review of Science and Technology. <https://www.researchgate.net/publication/367163451>

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). *Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review*. Journal of Affective Disorders, 227, 483–493. <https://doi.org/10.1016/j.jad.2017.11.048>

Yang, S. J. H., Lu, O. H. T., & Huang, A. Y. Q. (2018). *Predicting students' academic performance using multiple linear regression and principal component analysis*. Journal of Information Processing, 26, 170–179. <https://www.jstage.jst.go.jp/article/ipsjjip/26/0/26_170/_pdf>

Yesugade, K. D., Patil, K. V., & Naikwadi, K. B. (2024). *Regression-based machine learning models to predict student performance*. Journal of Engineering Education and Technology. <https://journaleet.in/download-article.php?Article_Unique_Id=JPR2202>