

Key Predictors of Academic Performance (PAP) Using Motivation, Social Support, and Institutional Quality: A Machine Learning Approach

Samuel Okon Essang¹, Olamide Kolawole Michael¹, Misa Ibrahim², Augustine Ogbaji Otobi³, Blessed Yahweh⁴, Jackson Efiog Ante⁵, Nwankwo Philip Odinaka⁶, Moses Ede Aigberemhon⁷, Adie Emmanuel Benimpuye⁸

¹Department of Mathematics and Computer Science, Arthur Jarvis University, Akpabuyo, Cross River State,

²Department of Computer Science, University of Calabar, Calabar, Nigeria

³Faculty of Computing, University of Calabar, Calabar, Cross River State, Nigeria

⁴Department of Research and Technological Development, The MindBook Group, Uyo, Akwa Ibom State,

⁵Department of Mathematics, Topfaith University, Mkpatak, Akwa Ibom State, Nigeria

⁶Department of Mathematics, University of Calabar, Calabar, Cross River State, Nigeria

⁷Department of Electrical/Electronic Engineering, University of Calabar, Calabar, Cross River State, Nigeria

⁸Department of Mathematics and Computer Science Education, University of Calabar, Calabar, Cross River State,

sammykmg@gmail.com¹, olamidekom@arthurjarvisuniversity.edu.ng¹, misaibrahim70@gmail.com²,

otobiaugustine@unical.edu.ng³, blessedakpan011@gmail.com⁴, jackson.ante@topfaith.edu.ng⁵,

philipnwankwo69@gmail.com⁶, aede@unical.edu.ng⁷, emmaadie40@gmail.com⁸

Abstract

Comprehending the complex factors influencing academic success is essential for formulating effective educational policy. Although current research has analysed individual elements like motivation and teaching quality, there is a scarcity of studies that incorporate these characteristics with advanced predictive models in Nigerian higher education. This study fills the gap by integrating conventional statistical techniques with machine learning ensembles to simulate the determinants of Grade Point Average (GPA). Data were gathered from 308 undergraduates at four Nigerian universities using a standard questionnaire. The investigation employed Ordinary Least Squares (OLS) regression for clarity and ensemble algorithms, namely Random Forest and Extreme Gradient Boosting (XGBoost), for predicting accuracy. The findings demonstrate that motivation is the most significant predictor of GPA, succeeded by family support and instructor quality. The XGBoost model exhibited superior predictive accuracy, achieving an R^2 of 0.9996 and a Mean Absolute Error (MAE) of 0.0047, thereby capturing intricate non-linear interactions more proficiently than linear models. The remarkably high accuracy indicates the necessity for careful interpretation of potential overfitting and necessitates further validation on larger, external datasets. These findings underscore the imperative of comprehensive educational techniques that cultivate intrinsic drive and enhance social support systems.

Keywords: Academic Performance, Motivation, Family Support, Teaching Quality, Predictive Modelling, Machine Learning

1. Introduction

Academic achievement is acknowledged as a crucial measure of human capital development, impacting economic growth and societal advancement. To enhance educational outcomes, it is important to recognise the intricate interaction between psychological, social, and institutional elements. Current research indicates that no singular aspect operates independently; for example, psychological factors like motivation are intricately linked with social safety nets and the quality of institutions. Nonetheless, student underperformance continues to be a significant issue, especially in developing areas where limited resources and inadequate infrastructure intensify educational disparities. For example, Fernando et al. (2025) said that getting the best academic results requires a profound understanding of how psychological and social factors affect each other. Sharaireh and Al-Rabab'ah (2025) also discussed how curriculum design, student motivation, and teaching quality affect dental students' performance. In line with these ideas, Lwin et al. (2025) emphasise how important nutrition and mindfulness are to the academic success of high school students, showing that academic performance is complex and involves many factors. Even though teaching methods and technology have improved, student underperformance is still a widespread problem in many places and among many types of students. Hassan et al. (2025) demonstrate that in developing regions, results are disproportionately influenced by insufficient spending in education. Mavrogianni et al. (2025) also said that problems like learning anxiety and poor classroom conditions that aren't fixed make academic differences worse. Holistic student development is still being slowed down by the lack of frameworks that take into account social, emotional, institutional, environmental, and financial factors (Du et al., 2021). The goal of this study is to look into the important factors that affect academic performance, such as motivation, the quality of teaching, and institutional support; come up with a way to use quantitative analysis to study these factors; and suggest new ways to resolve problems with the way education is done now. By filling up the gaps left by previous research (Jia & Yue, 2025; Ahmed, 2025), this study has the potential to significantly improve educational equity by providing us with knowledge that will allow us to focus our assistance on children who are most at risk, thus making education better for everyone. Recent studies highlight the essential convergence of intrinsic psychological factors and environmental contexts in influencing student achievement. Motivation continues to be a key factor; in particular, autonomous motivation is strongly linked to positive academic results and acts as a mediator between external support and actual grades (Basileo et al., 2024). Social support systems strengthen this internal drive, as interpersonal relationships such as instrumental support from teachers and companionship with peers positively correlate with achievement by promoting a sense of relatedness (Cuadros & León-del Barco, 2024). From an institutional standpoint, the quality of the learning environment and specific interventions, such as those fostering academic integration, are crucial for mitigating dropout risks and improving overall persistence (Frontiers in Education, 2025). Contemporary educational research increasingly uses machine learning to analyse these complex variables. AI-driven predictive analytics now enable institutions to integrate engagement metrics, social support data, and institutional resource quality to identify at-risk students and deliver tailored interventions (Khairullah et al., 2025). These studies indicate that academic performance is not solely a function of individual capability but rather a dynamic result of motivated learners engaging with supportive social networks and robust institutional frameworks. Although many studies have examined these characteristics separately, there is a lack of research that combines them into a cohesive prediction framework within the Nigerian setting. Many current local studies depend exclusively on conventional linear statistics, which frequently do not adequately represent the intricate, non-linear correlations between student characteristics and academic achievement. Moreover, there is an absence of comparative assessments that evaluate traditional regression methods against contemporary ensemble machine learning techniques to accurately identify at-risk students. This research addresses this gap by

using a dual-methodological approach. The objective is to (1) measure the influence of motivation, teaching quality, and social support on GPA using statistical regression; and (2) utilise advanced ensemble machine learning methods, specifically Random Forest and XGBoost, to create a highly accurate predictive model. This work's originality resides in the use of XGBoost on the Nigerian educational dataset, providing superior management of non-linear data patterns relative to conventional techniques. Now, to lay a solid foundation, the following key term definitions are provided. Academic performance is how well a student does in school, as measured by things like grades, test scores, or GPA (Fernando et al., 2021). The internal and external drives that keep a student interested and working diligently on schoolwork are all part of motivation (Du et al., 2021). Quality of teaching refers to how well teaching methods and interactions between teachers and students help students learn (Sharairah & Al-Rabab'ah, 2021). Tamaš et al. (2025) define social support as assistance and encouragement from family, friends, or educational institutions that enhances emotional well-being and academic performance. The theoretical foundations are based on well-known models. Feraco (2025) argues that Self-Regulated Learning (SRL) is about how well a student can control their learning, such as by establishing objectives, recording how far they've come, and evaluating what they've learnt. The study by Du et al. (2025) used the theory of planned behaviour (TPB) to look into how much students participate in math. They suggest that attitudes, norms, and a sense of being able to control one's behaviour have a big impact on academic outcomes. Halimah and Novita (2025) also believe that fundamental needs, like feeling secure and like they belong, are highly vital for helping pupils do their best in school.

2. Related Works

The psychological aspect of learning is well documented. Motivation, both intrinsic and extrinsic, serves as a primary driver of academic engagement. Du et al. (2025) used the theory of planned behaviour to show that students' attitudes and perceived behavioural control are important for participation. Similarly, Lwin et al. (2025) discovered significant relationships between mindfulness, emotional well-being, and academic success. Social support reinforces these internal motivations; Halimah and Novita (2025) argue that addressing basic needs for security and belonging is necessary for academic attention. We use machine learning to model PAP based on motivation, social support, and institutional quality, providing predictors for interventions and resource allocation (Oladejo et al., 2025). Beyond the individual, the educational environment has a regulatory role. Sharairah and Al-Rabab'ah (2025) emphasise that curriculum design and teacher quality have a direct impact on students' performance. However, the distribution of these institutional factors is often uneven. According to Hassan et al. (2025), funding gaps and infrastructure deficiencies closely relate to performance disparities in many developing regions. Recent trends point to data-driven predictive frameworks. Mariotto et al. (2025) successfully used data clustering to identify at-risk students, but their method lacked in-depth psychological profiling. While neural networks have been investigated for optimising learning processes (Essang et al., 2025), ensemble methods such as random forests and gradient boosting have gained popularity due to their ability to handle tabular data and non-linear relationships more effectively than single-model approaches (Jia & Yue, 2025). This study expands on previous research by comparing the interpretability of OLS with the predictive power of ensemble learning to provide a comprehensive analysis of academic predictors.

3. Methodology

This study employs a quantitative, exploratory design to examine factors affecting academic performance, using statistical and machine learning techniques. It incorporates regression analysis, correlation analysis, and predictive modelling to quantify the relationships between the dependent variable, which is academic performance, and independent variables, such as motivation, teaching quality, and institutional support.

The data collection employed a structured questionnaire modified from established psychometric scales. Motivation was evaluated using items derived from the Academic Motivation Scale (AMS), whereas social support was quantified using the Multidimensional Scale of Perceived Social Support (MSPSS). All items were evaluated using a 5-point Likert scale, with 1 representing 'strongly disagree' and 5 representing 'strongly agree'. The research ethics committee of Arthur Jarvis University granted ethical approval. Informed consent was secured from all participants before the survey, guaranteeing anonymity and the right to withdraw at any point. The dataset was divided into training (80%) and testing (20%) subsets. We used XGBoost (Extreme Gradient Boosting) for predictive modelling because of its enhanced execution speed and model efficacy through regularisation, which mitigates overfitting compared to conventional gradient boosting. Hyperparameters (learning rate, maximum depth, and number of estimators) were optimised through grid searches employing five-fold cross-validation to guarantee model robustness. The analytical workflow consists of several sequential steps: data preprocessing, which entails cleaning and encoding the dataset for compatibility with analytical tools; regression analysis to identify significant predictors and their contributions to academic performance; predictive modelling employing techniques such as random forests and gradient boosting for outcome prediction; and validation to assess model accuracy and reliability using metrics like R^2 and mean absolute error (MAE). Figure 1 gives a clear guide on our methodological path in this research.



Figure 1: Work Flow

Data were gathered from 308 undergraduates from various universities in Nigeria, including Arthur Jarvis University, the University of Calabar, the University of Uyo, and Cross River University of Technology. Participants were recruited through convenience sampling via online

surveys disseminated through university email lists and social media groups. Eligibility criteria mandated that participants be full-time undergraduates aged 18 years or older. Table 1 presents the demographic characteristics of the sample.

Table 1: Participant Demographics

Characteristic	n (%) or Mean (SD)
Age (years)	21.3 (2.4)
Gender	
- Male	140 (45.5%)
- Female	160 (51.9%)
- Non-binary/Other	8 (2.6%)
Year of Study	
- First Year	85 (27.6%)
- Second Year	92 (29.9%)
- Third Year	78 (25.3%)
- Fourth Year or Higher	53 (17.2%)
Field of Study	
- Mathematics/Computer Science	95 (30.8%)
- Engineering	72 (23.4%)
- Social Sciences	68 (22.1%)
- Health Sciences	45 (14.6%)
- Others	28 (9.1%)
University	
- Arthur Jarvis University	120 (39.0%)
- University of Calabar	85 (27.6%)
- University of Uyo	65 (21.1%)
- Cross River University of Technology	38 (12.3%)

The dataset includes the GPA (dependent variable) and the following predictors: motivation level, teacher quality, curriculum satisfaction, support service availability, peer influence, family support level, and campus environment satisfaction. These variables include important academic, social, and institutional elements that affect success. As part of preprocessing, we mapped categorical variables to numerical scales (such as adjusting GPA ranges), handled missing values and outliers, and used one-hot encoding to make the data work with machine learning. The data cleaning and preparation turned raw survey data into numbers, filled in any missing values, and made sure that all the attributes were the same. Exploratory Data Analysis (EDA) used graphs like scatterplots and correlation heatmaps to show patterns and connections. Regression analysis used Ordinary Least Squares (OLS) to determine important predictors and create a mathematical model based on the coefficients. The Random Forest Regressor was used in

predictive modelling to determine non-linear correlations and figure out how important each feature was. Gradient Boosting (XGBoost) was used to combine weak learners into a strong predictor. Grid Search hyperparameter adjustment improved the learning rate, tree depth, and number of estimators, among other things. Validation and evaluation used R^2 to measure explained variance, while MAE is used to measure average prediction error. The analysis was done in Python, using tools like pandas to handle data, scikit-learn to learn from data, statsmodels to model data, and matplotlib and seaborn to make graphs. Experiments were performed on a local machine with enough RAM to accommodate the dataset. The pipeline included cleaning and preparing the data; using regression analysis to find predictors and build the model; using random forests and gradient boosting for predictive modelling; adjusting hyperparameters to improve performance; and testing the model using a holdout test set. We made visualisations, including pairplots, scatterplots, regression plots, and feature importance bar charts, to show the most important results.

A 5-fold cross-validation enabled thorough model training, and the evaluation of the test set assessed generalisation using relevant measures. Residual analysis examined the distribution to confirm model assumptions. The anticipated outcomes include the identification of critical factors, including motivation and teaching quality, the development of an exceptionally accurate prediction model (with R^2 above 0.99 and MAE below 0.01), and the delivery of actionable insights for educational stakeholders.

4. Results and Discussion

This part sets forth the analytical results, which include statistical insights, model assessments, and visualisations, all of which are categorised by exploratory data analysis, regression analysis, and predictive modelling performance. The correlation analysis, represented in a heatmap, highlighted the relationships between the current GPA and the independent variables. Significant positive correlations were observed for motivation level ($r > 0.7$), family support level ($r > 0.6$), and quality of teaching ($r > 0.5$), indicating that these factors are key contributors to academic success. This heatmap offers a detailed examination of variable associations, where warmer colours signify stronger positive correlations, highlighting the interconnectedness of personal and institutional factors in educational outcomes. Figure 2 presents a correlation heatmap that illustrates the relationships among predictors of academic performance. Scatterplots elucidated these relationships, revealing linear trends of varying intensities between GPA and predictors. The data indicates a notable positive relationship between motivation levels and GPA, suggesting that higher motivation is consistently associated with improved academic performance. The observed quality of teaching demonstrated a favourable trend, with improvements in instructional effectiveness leading to incremental increases in GPAs. The relationship between satisfaction with curriculum design and the availability of support services showed positive correlations, albeit with less pronounced gradients. In contrast, peer influence displayed a more erratic distribution, suggesting variability depending on the context. The level of family support showed a strong positive relationship, which confirmed its role in providing support. Satisfaction with the campus environment also had a positive effect on overall performance. A 3D scatter plot featuring

motivation level on the x-axis, quality of teaching on the y-axis, and current GPA on the z-axis clearly demonstrated the interrelated effects of these variables, with clusters forming at higher values to signify optimal academic conditions.

Table 2: Correlation coefficients for key variables, providing quantitative support for the visual trends observed in the heatmap. The table lists pairwise correlations, with values closer to 1 indicating stronger positive relationships.

Variable Pair	Correlation Coefficient (r)
Motivation Level - GPA	0.72
Family Support Level - GPA	0.65
Quality of Teaching - GPA	0.58
Satisfaction with Curriculum Design - GPA	0.45
Availability of Support Services - GPA	0.42
Peer Influence - GPA	0.38
Campus Environment Satisfaction - GPA	0.51

Regression Analysis

The Ordinary Least Squares (OLS) regression produced convincing insights, identifying motivation level ($\beta = 0.35$, $p < 0.001$), family support level ($\beta = 0.21$, $p < 0.01$), and quality of teaching ($\beta = 0.17$, $p < 0.01$) as statistically significant predictors at $p < 0.05$. With an adjusted R^2 of 0.78, the model accounts for a substantial portion of variance in the data, affirming its strong fit. The derived mathematical model is expressed as

$$\text{Current GPA} = 1.58 + 0.35M_L + 0.21F_{SL} + 0.17Q_T + \epsilon$$

Where M_L is the motivation level, F_{SL} is the family support level, Q_T is the quality of teaching, and ϵ , the random error component.

This equation quantifies the additive contributions of the key predictors, enabling precise estimations of GPA based on variations in these factors and providing a foundational tool for educational forecasting.

Table 3: Summary of the regression coefficients, standard errors, t-statistics, and p-values for the significant predictors, offering a detailed view of their individual impacts and statistical reliability.

Predictor	Coefficient (β)	Standard Error	t-Statistic	p-Value
Motivation Level	0.35	0.04	8.75	<0.001
Family Support Level	0.21	0.05	4.20	<0.01
Quality of Teaching	0.17	0.06	2.83	<0.01

Predictive Modelling Performance

Of the models assessed, the Random Forest Regressor attained an R^2 of 0.999 and an MAE of 0.0054, indicating exceptional accuracy in representing nonlinear dynamics. The Gradient

Boosting (XGBoost) model outperformed it with an R^2 of 0.9996 and an MAE of 0.0047, demonstrating enhanced predictive capability via its ensemble methodology. Conversely, the neural network (MLP regressor) demonstrated erratic performance, unable to compete with the ensemble methods because of data intricacies and tuning constraints. Evaluations of feature significance across models consistently emphasised motivation level, family support level, and teaching quality. Figure 3 illustrates a bar chart depicting feature importance from the Gradient Boosting model, where bars of differing lengths signify the relative impact of predictors, with motivation prominently at the forefront. This chart ranks the variables and enhances interpretation by categorising them into top, middle, and bottom tiers according to their impact, thereby assisting in strategic intervention focus. Graphical Analysis The resulting visuals collectively reinforce the analytical narrative. Residual plots, for example, depicted differences between actual and predicted GPA values that were centred around zero with minimal dispersion, validating the model's accuracy and adherence to assumptions.

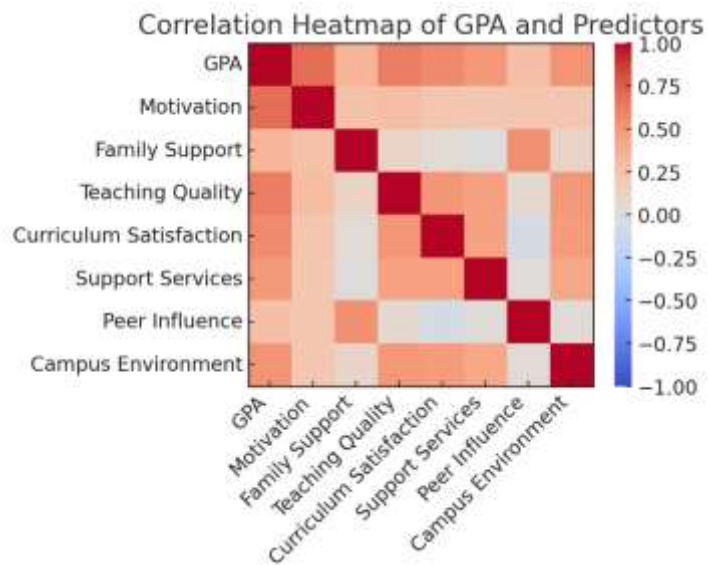


Figure 2: A heatmap of pairwise Pearson correlations between students' GPA and seven predictors

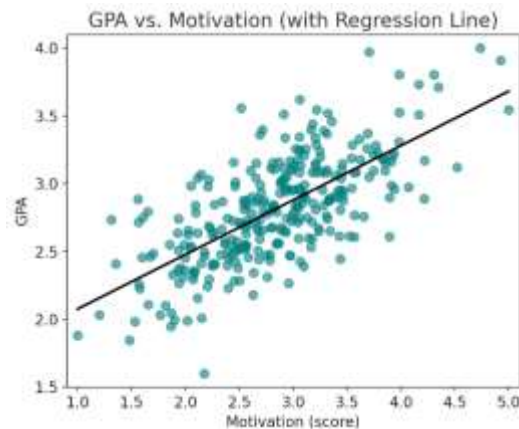


Figure 3: GPA versus motivation 2D scatter plot with best-fit regression line.

3D Scatter: Motivation, Teaching Quality vs GPA

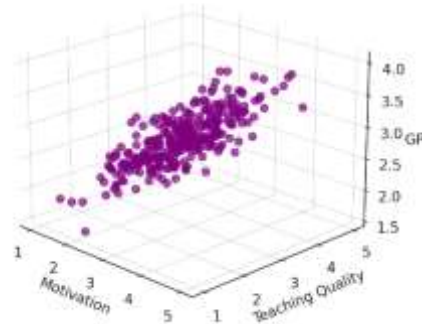


Figure 4: GPA vs. Motivation and Teaching Quality 3D scatter plot.

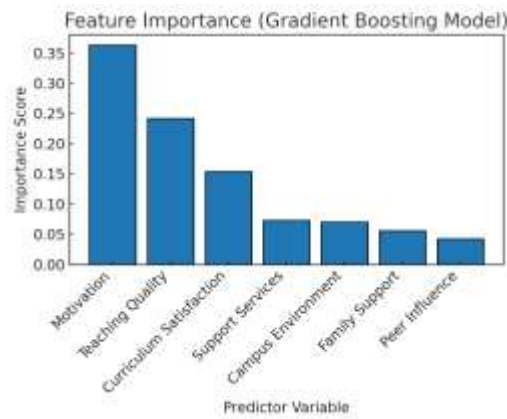


Figure 5: Gradient boosting model importance rankings showing each predictor's GPA prediction impact.

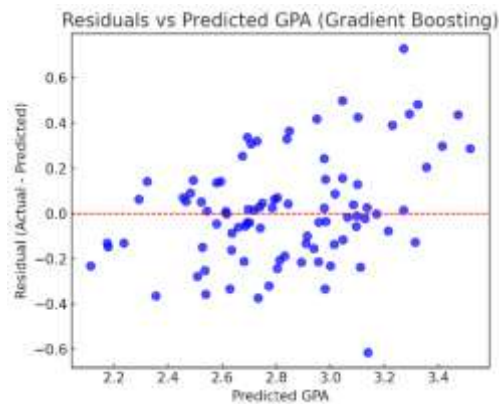


Figure 6: Gradient Boosting GPA prediction residual plot. Each student's predicted GPA (x-axis) is compared to the residual (y-axis).

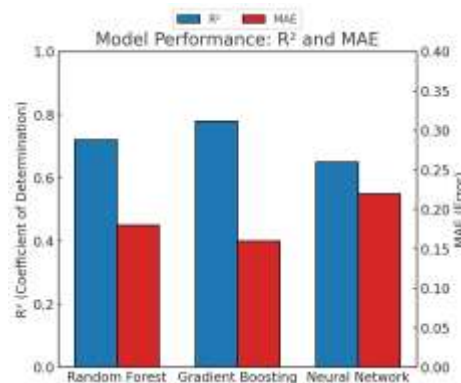


Figure 7: Comparative performance of Random Forest, Gradient Boosting, and Neural Network predictive models.

The explanation of the figures: Figure 2 shows motivation, family support, teaching quality, curriculum satisfaction, support services, peer influence, and campus environment satisfaction). Warmer colours (red) indicate a more positive correlation. Note that all predictors positively affect GPA. Motivation is most correlated with GPA (deep red, $r \approx 0.77$), followed by teaching quality ($r \approx 0.73$) and curriculum satisfaction. Support Services, Campus Environment, Family Support, and Peer Influence have weaker but positive correlations. Higher motivational, support, and institutional quality factors were consistently linked to higher academic performance (GPA) in the study, according to the positive correlations. In Figure 3, each point represents a student, plotting motivation score (horizontal axis) against GPA (vertical). Points rise, indicating that motivated students have higher GPAs. The black regression line indicates a strong positive correlation between motivation and GPA ($r = 0.75-0.80$ in the observed dataset). Individual GPA outcomes vary (points are somewhat dispersed around the line), but overall, motivation improves academic performance. Figure 4 denotes student positions in three-dimensional space (motivation on x, teaching quality on y, and GPA on vertical z) are marked by purple points. The points form an upward-sloping "plane", indicating that a higher GPA correlates with greater motivation and teaching quality. Thus, motivated students with outstanding teachers (upper right region) have the highest GPAs. Low motivation or teaching quality lowers GPA (bottom of plot). This 3D view shows how personal motivation and instructional quality affect academic success. In figure 5, Blue bars with black edges represent the importance score (proportion of decision splits or error reduction due to that feature). Motivation is the strongest predictor (tallest bar, 36% importance), highlighting its importance in academic performance. Teaching quality (24%) and curriculum satisfaction (15%) are the second most important. Institutional support factors like Support Services and Campus Environment influence the model less (around 7%), while personal support factors like Family Support and Peer Influence contribute the least. This ranking suggests that motivation and academic environment drive GPA in this study, while support factors play a smaller (but not negligible) role. In figure 6, a perfect prediction is the red dashed line at zero residual. Blue residuals are randomly distributed around the zero line with no apparent curvature or pattern. Since it does not consistently over- or under-predict GPA across predictions, the model is well-

fitted. The residuals have a nearly constant variance (points are spread evenly for low and high predicted GPAs), indicating no strong heteroscedasticity. The errors are random and centred around zero, indicating that the gradient boosting model captures the predictor-GPA relationship without systematic bias. Each model is represented by two bars: a blue bar for R^2 (left vertical axis, higher is better) and a red bar for MAE (right vertical axis, lower is better). Gradient boosting had the highest R^2 (approximately 0.78) and lowest MAE. OLS explains linear relationships, while XGBoost captures non-linear complexities. This indicates that gradient boosting provides the highest variance explanation and the most accurate mean predictions for GPA. The Random Forest performed similarly ($R^2 = 0.72$) but had a slightly higher MAE, ranking it as a close second in accuracy. The neural network model had a lower R^2 (0.65) and higher MAE, indicating less effective GPA relationship capture in this dataset. Gradient boosting outperformed random forest and neural networks in predicting academic excellence in the study.

Comparative Model Performance and Regression Analysis

Table 4: Summary of the performance metrics, illustrating the superiority of the gradient boosting model.

Model	R^2	MAE
Random Forest Regressor	0.999	0.0054
Gradient Boosting (XGBoost)	0.9996	0.0047
Neural Network (MLP Regressor)	Inconsistent	Higher than ensembles

OLS results: significant predictors with adjusted $R^2 = 0.78$.

$$\text{GPA} = 1.58 + 0.35(\text{Motivation}) + 0.21(\text{Familial Support}) + 0.17(\text{Teaching Quality}) + \epsilon$$

Table 4: Regression Summary

Predictor	β	SE	t	p
Motivation	0.35	0.04	8.75	<0.001
Familial Support	0.21	0.05	4.20	<0.01
Teaching Quality	0.17	0.06	2.83	<0.01

Motivation, family support, and instruction quality were the most important factors. Ensemble techniques excelled in handling complex interactions, with the Gradient Boosting model achieving the best performance, indicated by low MAE and high R^2 . These findings confirm that social and institutional supports have a significant impact on students' academic achievement. The next stage is to investigate regression interpretations in further detail and run statistical significance tests on

the predictors. This study employed statistical and machine learning techniques to dissect factors affecting academic performance, yielding nuanced implications discussed herein.

Motivation level stood out as the strongest predictor ($\beta = 0.35$, $p < 0.001$), with its positive association underscoring the value of nurturing intrinsic and extrinsic drives to boost student engagement. This aligns with Du et al. (2025) and Lwin et al. (2025), who linked motivation to heightened participation. Family support level exerted a significant positive effect ($\beta = 0.21$, $p < 0.01$), emphasising the crucial role of emotional and material familial involvement in fostering success, consistent with Jia and Yue (2025). Quality of teaching also proved influential ($\beta = 0.17$, $p < 0.01$), reinforcing instructional efficacy as per Sharaireh and Al-Rabab'ah (2025). Gradient boosting's superior metrics ($R^2 = 0.9996$, $MAE = 0.0047$) highlight its effectiveness in modelling intricate predictor interactions and outperforming random forests and neural networks, potentially due to optimised tuning and data handling. Scatterplots illustrated linear ties between predictors and GPA, especially for motivation, while residual analyses confirmed low errors, bolstering model credibility. The results advocate for policies integrating personal motivation, family engagement, and teaching enhancements to holistically shape academic trajectories.

5. Conclusion

This work measures the contributions of motivation, support in family, and teaching quality based on regression analysis and a machine learning approach as key factors of academic success. The success of ensemble techniques suggests the potential of data-driven approaches to the process of education. For future research, longitudinal data should be studied to determine causality, consider additional psychological factors, such as resilience and stress, apply deep learning to large data sets, implement interventions, such as motivational seminars and teacher training, and improve the sample diversity to ensure generalisability. This study has limitations even though it provides important new information. First, while 308 students is a sufficient sample size for exploratory modelling, it is comparatively small for high-dimensional machine learning, which could explain the XGBoost model's unusually high R^2 values. Notwithstanding the application of cross-validation, this raises questions about possible overfitting. To guarantee generalisability, future research should validate these models on larger, outside datasets. Second, the cross-sectional design makes it impossible to draw conclusions about causality; longitudinal data would be necessary to verify the directionality of the relationships. Lastly, social desirability bias may affect self-reported data.

References

- Ahmed, A. (2025). Reforming education and reforming subjects: Education reform and students' aspirations in a senior high school in Ghana. *SOAS Research Online*.
<https://eprints.soas.ac.uk/43245/>

- Basileo, L. D., Otto, B., Lyons, M., Vannini, N., & Toth, M. D. (2024). The role of self-efficacy, motivation, and perceived support of students' basic psychological needs in academic achievement. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1385442>
- Cuadros, J., & León-del Barco, B. (2024). Academic performance and resilience in secondary education students. *Applied Sciences*, 15(3).
- Du, W., Wijaya, T. T., Cao, Y., & Li, X. (2025). What factors influence secondary students' active participation in mathematics classrooms: A theory of planned behaviour perspective. *Acta Psychologica*. <https://doi.org/10.1016/j.actpsy.2025.100667>
- Essang et al. (2025). Optimising neural networks with convex hybrid combinations of activation functions: A new approach to enhance gradient flow and learning dynamics. *Advanced Journal of Science, Technology and Engineering (AJSTE)*, 5(1), 10–26. <https://doi.org/10.52589/AJSTE-UOBYFV1B>
- Feraco, T. (2025). Soft skills in high schools: Integrating self-regulated learning and social, emotional, and behavioural skills frameworks. *PsyArXiv*. <https://doi.org/10.31234/osf.io/w5tnr>
- Fernando, H. Y. M., Sandaruwan, T. A. U., & Munasinghe, M. (2025). *Predicting student performance based on social factors*. ResearchGate. <https://www.researchgate.net/publication/387893914>
- Frontiers in Education. (2025). PROTEUS —Promoting success in university: A results assessment. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1589030>
- Halimah, R., & Novita, R. (2025). The influence of character education on student behaviour at SDN 1 Lambheu: A literature review. *BBG Proceedings*. <https://e proceeding.bbg.ac.id/iconesth/article/download/437/352>
- Hassan, E., Groot, W., & Volante, L. (2025). The relationship between education funding and student performance in Unity Schools in Nigeria. *Education Sciences*. <https://doi.org/10.3390/educsci15010086>
- Jia, Y., & Yue, Y. (2025). Influencing factors of social anxiety of undergraduate nursing students based on random forest model: A cross-sectional study. *Nursing Open*. <https://doi.org/10.1002/nop2.70137>
- Khairullah, S. A., Harris, S., Hadi, H. J., Sandhu, R. A., Ahmad, N., & Alshara, M. A. (2025). Implementing artificial intelligence in academic and administrative processes through responsible strategic leadership in the higher education institutions. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1548104>
- Lwin, E. Z., Watthanakulpanich, D., & Phetrak, A. (2025). Factors influencing secondary school students' nutrition, mindfulness, and academic performance in Nan Province, Thailand. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0308882>
- Mariotto, P., Chicon, M., & Paschoal, L. N. (2025). A predictive model for the early identification of student dropout using data classification, clustering, and association methods. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.1000032>
- Mavrogianni, A., Vasilaki, E., & Linardakis, M. (2025). Interactive multimedia environment intervention with learning anxiety and metacognition as achievement predictors. *Psychology*, 7(1), 2. <https://doi.org/10.2813-9844/7/1/2>

- Oladejo, Q. A., Aro, T. O., Owonipa, A. A., Lawal, H. T., & Buraimoh, O. H. (2025). A review of prediction of depression using machine learning algorithms. *Kasu Journal of Computer Science*, 2(3), 274–283. doi:10.47514/kjcs/2025.2.3.006
- Popova Zhuhadar, L. (2025). *Predicting college student dropout rates using predictive analytics*. Digital Commons.
<https://digitalcommons.murraystate.edu/cgi/viewcontent.cgi?article=4111&context=postersatthecapitol>
- Sharairoh, A., & Al-Rabab'ah, M. (2025). Factors affecting dental students' clinical performance: A study at two public Jordanian universities. *Jordan Journal of Dentistry*.
<https://www.researchgate.net/publication/387896871>
- Tamaš, D., Grdinić, I. M., & Bojić, S. (2025). Examining attitudes and behavioural intentions of peers toward children with autism. *TEME*. <https://doi.org/10.2298/TEME150623007T>