COMPARATIVE BAYESIAN ANALYSIS OF FACTORS AFFECTING STUDENT PERFORMANCE

Dhamini Rao Pantra, Eshani Nandy, Mishaal Khan, Munia Humaira

University of Waterloo

1. INTRODUCTION

Educational institutions generate vast amounts of data through student information systems (SIS), learning management systems (LMS), and other digital platforms. These datasets hold immense potential for improving educational strategies and decision-making processes to enhance student performance. In this project, we use a dataset from an undergraduate student survey collected in 2024 from the Universiti of Malaya [1] and employ a Bayesian statistical framework implemented in Stan to investigate the specific academic and non-academic factors that determine a student's CGPA in higher education and to explore how these factors interact.

Our primary research question is: **What specific academic and non-academic factors influence a student's CGPA in higher education, and how do these factors interact?**

To address this question, we propose the following specific hypotheses:

- **H1**: Students with a higher CGPA last semester, good preparation, and higher attendance will have significantly higher final CGPA.
- **H2**: Socioeconomic factors like income and hometown do not have much impact on students in higher education.
- **H3**: Excessive gaming has a negative effect on CGPA, but this effect can be moderated by extracurricular involvement.
- **H4**: The benefits of study preparation and previous semester performance on CGPA depend on students' non-academic behaviors.

These hypotheses guide our analysis and are designed to test both established findings and novel insights regarding the interplay between academic history and lifestyle factors on student performance. The following sections detail our data preparation, methodological framework, results, and discussion, with the goal of informing personalized teaching strategies and targeted educational interventions.

2. METHODOLOGY

This section contains three parts; the first outlines the statistical summary of the dataset and the preprocessing steps. The subsequent parts from section 2.2 to 2.5 focus on the Bayesian modeling approach used to build the necessary models. The final part discusses the evaluation metrics employed to compare the models.

2.1. Data Summarization and Preprocessing

The dataset comprises 493 student records and includes 16 academic, socioeconomic, and behavioral attributes. Academic features include department, preparation time, attendance, and current semester as categorical columns and performance from the previous semester (Last), Higher Secondary score (HSC), and Secondary School score (SSC) as numerical data. The average number of semesters completed by the participants was 5.32 (range: 2–12). The non-academic features include gender, hometown, part-time job status, extracurricular participation, monthly family income, time spent gaming, and computer and English language proficiency, all of which are categorical data. Regarding demographics, 56.8% are from villages, 6.9% had part-time jobs, 41.6% participated in extracurricular activities, and 66.5% were male.

Scikit-learn's StandardScaler was used to standardize the continuous features and the target variable, CGPA. The categorical features with ordinal values underwent ordinal encoding with distinct integer values corresponding to their order, and the nominal features were one-hot encoded as binary columns $(1 : True, 0 : False)$. The Department feature, representing students' academic departments, had significantly varying sample sizes, with 443 students enrolled in the Department of Computer Science and Engineering (CSE) and the remaining 50 students spread across 9 different departments. Using the Department as a predictor with such disparity may introduce bias and limit the model's applicability. Section 2.2 provides further details on the decision to drop the Department predictor. We have performed an exploratory data analysis to get a sense of the correlations between variables in the dataset, which is shown as a heatmap in Figure 1. We can see that Last and Scaled CGPA are strongly posi-

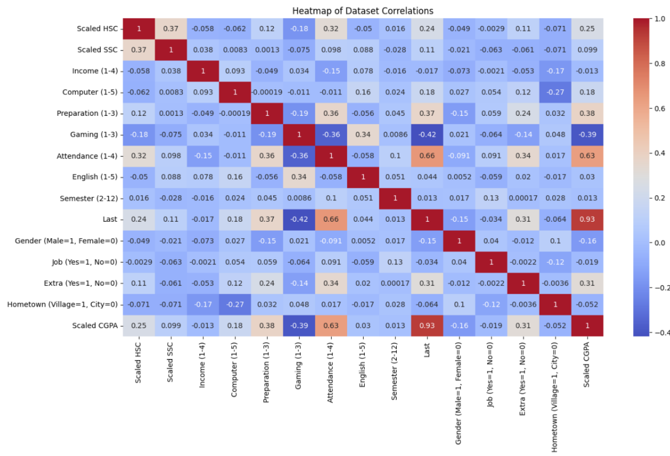

Fig. 1. Heatmap of correlations between the variables in the dataset.

tively correlated, with Attendance and Preparation also showing moderate positive correlations with scaled CGPA, while Gaming shows a moderate negative correlation. This gives us a sense of the predictors that will feature most prominently in our regression analysis.

2.2. Full-featured Model

Our full Bayesian multiple linear regression model includes all academic and non-academic factors as predictors. It is specified in the form:

$$\begin{aligned} y = \beta_0 &+ \beta_1 \cdot \text{HSC} + \beta_2 \cdot \text{SSC} + \beta_3 \cdot \text{Income} \\ &+ \beta_4 \cdot \text{Computer} + \beta_5 \cdot \text{Preparation} + \beta_6 \cdot \text{Gaming} \\ &+ \beta_7 \cdot \text{Attendance} + \beta_8 \cdot \text{English} + \beta_9 \cdot \text{Semester} \\ &+ \beta_{10} \cdot \text{Last} + \beta_{11} \cdot \text{Gender\_Male} + \beta_{12} \cdot \text{Job\_Yes} \\ &+ \beta_{13} \cdot \text{Extra\_Yes} + \beta_{14} \cdot \text{Hometown\_Village} + \varepsilon. \end{aligned} \tag{1}$$

here, y represents the scaled CGPA for a given student, and each β corresponds to the effect of a specific predictor:

- β_0 is the intercept.
- β_1 through β_{14} represent the coefficients for the predictors.
- ε is the error term assumed to be normally distributed with standard deviation σ.

The model is based on the assumption that each observed CGPA is normally distributed around a linear combination of these predictors, i.e.,

$$Y_i \sim \mathcal{N}(\alpha + \mathbf{X}_i \cdot \boldsymbol{\beta}, \sigma), \tag{2}$$

where Y_i represents the scaled CGPA for the ith student, α is the intercept, $\mathbf{X}_i$ is the vector of predictors for that student, $\boldsymbol{\beta}$ is the vector of regression coefficients, and σ denotes the standard deviation of the residual error. We adopted diffuse, weakly informative priors *(Normal(0, 10))* for the intercept, coefficients, and error term to allow the data to drive inference. This approach aligns well with the Bayesian principle of integrating prior knowledge with observed data, ensuring robust and interpretable results [2].

An important consideration during data preprocessing was the handling of the Department column. Initially, this predictor was included in the model; however, the severe imbalance in the number of students across departments led to convergence issues (with $\hat{R} \approx 3$). We then experimented with grouping departments into a binary variable (CSE vs. non-CSE), but this alternative did not improve the model fit. Consequently, we dropped the Department predictor entirely from subsequent analyses to ensure reliable estimation.

A data dictionary was constructed for Stan, which included the number of observations (N), the number of predictors (K), the predictor matrix (X), and the vector of scaled CGPA values (Y). The Stan model code defines the data using these elements and specifies the parameters using α, β, and σ, along with the priors and the likelihood as described above. This full-featured model was implemented in Stan [3] via its Python interface, cmdstanpy. The model was compiled and sampled using the No-U-Turn Sampler (NUTS-HMC) with 4 chains, each running 1000 warm-up and 2000 sampling iterations, ensuring robust convergence and reliable posterior estimation.

2.3. Model - Academic Features

To assess the hypothesis H1 on student performance, we developed two Bayesian regression models using the same approach and diffuse priors as in the full-featured model. For the academic main effects model, we used all 6 academic predictors. The regression formula is:

$$\begin{aligned} y = \beta_0 &+ \beta_1 \cdot \text{HSC} + \beta_2 \cdot \text{SSC} + \beta_3 \cdot \text{Preparation} \\ &+ \beta_4 \cdot \text{Attendance} + \beta_5 \cdot \text{Semester} + \beta_6 \cdot \text{Last} + \varepsilon. \end{aligned} \tag{3}$$

This model captures the individual contributions of each academic factor to CGPA.

For the academic interaction model, we focus on a subset of academic predictors and their interactions to better understand how these factors jointly influence CGPA. Its regression formula is:

$$\begin{aligned} y = \beta_0 &+ \beta_1 \cdot \text{Preparation} + \beta_2 \cdot \text{Attendance} + \beta_3 \cdot \text{Last} \\ &+ \beta_4 \cdot (\text{Last} \times \text{Preparation}) + \beta_5 \cdot (\text{Preparation} \times \text{Attendance}) \\ &+ \beta_6 \cdot (\text{Last} \times \text{Attendance}) + \varepsilon. \end{aligned} \tag{4}$$

This interaction model helps reveal whether the effect of one academic predictor is contingent on the levels of another.

Together, these models allow us to compare the isolated effects of academic factors with their combined effects, providing a comprehensive understanding of their influence on student performance.

2.4. Model - Non-academic Factors

We also developed two regression models using non-academic factors to assess hypotheses H2 and H3 using the same approach and diffuse priors as in the full-featured model. The main effect model's regression formula is:

$$\begin{aligned} y = \beta_0 &+ \beta_1 \cdot \text{Gender\_Male} + \beta_2 \cdot \text{Income} + \beta_3 \cdot \text{Computer\_Yes} \\ &+ \beta_4 \cdot \text{Gaming} + \beta_5 \cdot \text{Job\_Yes} + \beta_6 \cdot \text{English} \\ &+ \beta_7 \cdot \text{Extra\_Yes} + \beta_8 \cdot \text{Hometown\_Village} + \varepsilon. \end{aligned} \tag{5}$$

The second model incorporated interaction terms to capture the combined effects of multiple non-academic factors. The regression formula for this model is:

$$\begin{aligned} y = \beta_0 &+ \beta_1 \cdot \text{Gender\_Male} + \beta_2 \cdot \text{Computer\_Yes} \\ &+ \beta_3 \cdot \text{Gaming} + \beta_4 \cdot \text{Extra\_Yes} + \beta_5 \cdot \text{English} \\ &+ \beta_6 \cdot (\text{Gender\_Male} \times \text{Gaming}) \\ &+ \beta_7 \cdot (\text{Computer\_Yes} \times \text{Gaming}) \\ &+ \beta_8 \cdot (\text{Gaming} \times \text{Extra\_Yes}) \\ &+ \beta_9 \cdot (\text{Computer\_Yes} \times \text{Extra\_Yes}) \\ &+ \beta_{10} \cdot (\text{Gender\_Male} \times \text{Extra\_Yes}) \\ &+ \beta_{11} \cdot (\text{Gender\_Male} \times \text{English}) + \varepsilon. \end{aligned} \tag{6}$$

2.5. Models - Interactive Academic and Non-academic Factors

For testing hypothesis H4, we developed iterative models where we start with a model that has only main effects of the significant academic and non-academic factors, with the subsequent models including interactions between these factors. The models are as follows:

1. Model Y_1 (Main Effects Only):

$$\begin{aligned} y_1 = \beta_0 &+ \beta_1 \cdot \text{Preparation} + \beta_2 \cdot \text{Last} + \beta_3 \cdot \text{Gender\_Male} \\ &+ \beta_4 \cdot \text{Computer} + \beta_5 \cdot \text{English} + \beta_6 \cdot \text{Gaming} \\ &+ \beta_7 \cdot \text{Extra\_Yes} + \varepsilon_1. \end{aligned} \tag{7}$$

2. Model Y_2 (Adding Preparation × Gender_Male):

$$\begin{aligned} y_2 = \beta_0 &+ \beta_1 \cdot \text{Preparation} + \beta_2 \cdot \text{Last} + \beta_3 \cdot \text{Gender\_Male} \\ &+ \beta_4 \cdot \text{Computer} + \beta_5 \cdot \text{English} + \beta_6 \cdot \text{Gaming} \\ &+ \beta_7 \cdot \text{Extra\_Yes} + \beta_8 \cdot (\text{Preparation} \times \text{Gender\_Male}) \\ &+ \varepsilon_2. \end{aligned} \tag{8}$$

3. Model Y_3 (Adding Preparation × Computer):

$$\begin{aligned} y_3 = \beta_0 &+ \beta_1 \cdot \text{Preparation} + \beta_2 \cdot \text{Last} + \beta_3 \cdot \text{Gender\_Male} \\ &+ \beta_4 \cdot \text{Computer} + \beta_5 \cdot \text{English} + \beta_6 \cdot \text{Gaming} \\ &+ \beta_7 \cdot \text{Extra\_Yes} + \beta_8 \cdot (\text{Preparation} \times \text{Gender\_Male}) \\ &+ \beta_9 \cdot (\text{Preparation} \times \text{Computer}) + \varepsilon_3. \end{aligned} \tag{9}$$

In a similar manner the other interaction terms are added iteratively.

4. Model Y_4 (Adding Preparation × English to y_3):

$$y_4 = y_3 + \beta_{10} \cdot (\text{Preparation} \times \text{English}) + \varepsilon_4. \tag{10}$$

5. Model Y_5 (Adding Last × Gaming to y_4):

$$y_5 = y_4 + \beta_{11} \cdot (\text{Last} \times \text{Gaming}) + \varepsilon_5. \tag{11}$$

6. Model Y_6 (Adding Last × Gender_Male to y_5):

$$y_6 = y_5 + \beta_{11} \cdot (\text{Last} \times \text{Gender\_Male}) + \varepsilon_6. \tag{12}$$

7. Model Y_7 (Adding Last × Extra_Yes to y_6):

$$y_7 = y_6 + \beta_{13} \cdot (\text{Last} \times \text{Extra\_Yes}) + \varepsilon_7. \tag{13}$$

8. Model Y_8 (Adding Last × English to y_7):

$$y_8 = y_7 + \beta_{14} \cdot (\text{Last} \times \text{English}) + \varepsilon_8. \tag{14}$$

9. Model Y_9 (Adding Last × Computer to y_8):

$$y_9 = y_8 + \beta_{15} \cdot (\text{Last} \times \text{Computer}) + \varepsilon_9. \tag{15}$$

This formulation links each model to its predictors and interaction terms, allowing us to sequentially assess the contribution of each interaction to the overall prediction of scaled CGPA. Similar to the full model, all these models were implemented in Stan using the same sampling settings and non-informative priors. The Stan files for these interaction models maintain the same structure as the full model, with the primary difference being the predictor variables.

2.6. Comparison Techniques

We employed a suite of quantitative metrics and a cross-validation technique to rigorously evaluate the performance of our regression models.

- We calculated the R^2 value to measure the variability in scaled CGPA explained by the predictors; a higher value indicates better explanatory power.
- We computed Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to balance model fit and complexity by penalizing additional parameters; a lower value suggests a more parsimonious model with strong predictive accuracy.

Parameter	Model Y_1	Model Y_2	Model Y_3	Model Y_4	Model Y_5	Model Y_6	Model Y_7	Model Y_8	Model Y_9
β_1	0.068 (0.029)	0.053 (0.043)	0.031 (0.074)	0.076 (0.116)	0.08 (0.117)	0.063 (0.117)	0.062 (0.117)	0.072 (0.119)	0.061 (0.12)
β_2	0.894 (0.022)	0.894 (0.021)	0.894 (0.021)	0.894 (0.021)	0.806 (0.092)	0.749 (0.097)	0.768 (0.11)	0.751 (0.117)	0.767 (0.121)
β_3	-0.05 (0.037)	-0.094 (0.101)	-0.088 (0.104)	-0.085 (0.104)	-0.09 (0.105)	-0.043 (0.105)	-0.041 (0.107)	-0.047 (0.108)	-0.041 (0.108)
β_4	0.012 (0.014)	0.012 (0.014)	-0.001 (0.037)	-0.008 (0.04)	-0.006 (0.04)	-0.01 (0.041)	-0.014 (0.04)	-0.013 (0.041)	-0.022 (0.044)
β_5	-0.01 (0.022)	-0.01 (0.022)	-0.009 (0.022)	0.017 (0.056)	0.019 (0.056)	0.005 (0.057)	0.007 (0.057)	0.012 (0.058)	0.013 (0.058)
β_6	0 (0.03)	0.00E+00 (0.03)	0.00E+00 (0.031)	0.001 (0.031)	-0.017 (0.035)	-0.021 (0.035)	-0.019 (0.036)	-0.017 (0.036)	-0.018 (0.036)
β_7	0.057 (0.037)	0.056 (0.037)	0.056 (0.037)	0.056 (0.038)	0.055 (0.037)	0.06 (0.038)	0.062 (0.038)	0.061 (0.037)	0.06 (0.038)
β_8		0.026 (0.055)	0.022 (0.057)	0.02 (0.057)	0.02 (0.057)	-0.012 (0.059)	-0.014 (0.06)	-0.01 (0.06)	-0.014 (0.06)
β_9			0.008 (0.02)	0.011 (0.022)	0.01 (0.022)	0.013 (0.022)	0.015 (0.022)	0.015 (0.022)	0.02 (0.024)
β_{10}				-0.015 (0.03)	-0.016 (0.03)	-0.01 (0.03)	-0.011 (0.03)	-0.014 (0.031)	-0.015 (0.031)
β_{11}					0.032 (0.032)	0.032 (0.032)	0.028 (0.035)	0.025 (0.035)	0.024 (0.035)
β_{12}						0.075 (0.043)	0.074 (0.042)	0.073 (0.043)	0.075 (0.043)
β_{13}							-0.017 (0.04)	-0.019 (0.04)	-0.019 (0.041)
β_{14}								0.008 (0.019)	0.012 (0.02)
β_{15}									-0.009 (0.016)
α	-0.105 (0.113)	-0.078 (0.122)	-0.044 (0.159)	-0.124 (0.223)	-0.076 (0.227)	-0.021 (0.230)	-0.024 (0.228)	-0.042 (0.233)	-0.018 (0.235)
R^2	0.859	0.859	0.860	0.860	0.860	0.861	0.861	0.861	0.861

Fig. 2. R^2 values and parameter estimates $(\beta_1 - \beta_{15})$ and intercept (α) for Models $Y_1 - Y_9$, with standard errors enclosed in parentheses. The findings show that the model fit varies and that the coefficients are stable across specifications.

- Leave-one-out cross-validation (LOO-CV) was used to assess generalizability by iteratively excluding each observation, fitting the model, and aggregating prediction errors to obtain a robust out-of-sample performance estimate.

A thorough evaluation of both predictive accuracy and model complexity is ensured by comparing R^2, AIC, BIC, and LOO-CV across all models. By balancing generalizability and goodness-of-fit, this multi-criteria evaluation helps find the most reliable model.

3. RESULTS

In this section, we present the summarized results of our models, as illustrated in Figures 2 and 3. We analysed the performance of each model, ultimately selecting the one that best predicts scaled CGPA to address our main research question. We also evaluate how the experimental findings support our secondary hypotheses.

	Full Model	Academic Model	Academic Interaction Model	Non-Academic Model	Non-Academic Interaction Model	Academic:Non-Academic Model
AIC	459.701	449.304	436.416	1253.660	1234.866	455.311
BIC	526.909	482.908	470.020	1295.666	1289.472	514.119
LOO	-237.07	-231.46	-225.44	-627.41	-618.15	-235.41
R^2	0.861	0.861	0.863	0.285	0.320	0.861

Fig. 3. Evaluation of model performance metrics (AIC, BIC, LOO, and R^2) for Full, Academic, Non-Academic, and Interaction models compared to other model specifications.

3.1. Full Model

Our full model results demonstrate that academic performance is the primary driver of scaled CGPA. In this model, the coefficient for Last (last semester CGPA) is very high ($\approx$ 0.88), which confirms that prior academic performance is the most significant predictor, which is aligned with [4]. This directly supports H1. In addition, Preparation shows a moderate positive effect with a coefficient of approximately 0.062, and other academic variables contribute to the overall model fit. With robust convergence indicated by an $\hat{R}$ value of approximately 1.00267 and an $R^2 \approx 0.861$ (86.1%), the model provides strong evidence that academic factors dominate the prediction of CGPA. These results answer our research question by highlighting that academic performance is the key determinant of student success in higher education.

3.2. Academic models (main and interaction effects)

The academic main effects model explains 86.1% of the variance in scaled CGPA. In this model, the coefficient for Last (last semester CGPA) is approximately 0.89, indicating a strong positive association with the final CGPA. Preparation and Attendance have smaller, yet positive, effects, with coefficients around 0.07 and 0.03, respectively, while HSC, SSC, and Semester contribute minimally. These magnitudes clearly show that a student's prior academic performance is the most influential predictor of their final outcome, which directly supports H1. When we extend the model to include interactions among academic predictors, the explained variance increases slightly to 86.3%. In this interaction model, the magnitude of the coefficient for Last increases to about 1.13, further emphasizing its dominant role. Interestingly, the main effects for Preparation and Attendance shift from positive to negative (approximately –0.1192 and –0.0745, respectively) when their interactions with Last are introduced. This change in sign suggests that the benefits of a high last semester CGPA may be moderated by the levels of preparation and attendance, revealing that the impact of these factors is more complex than when considered in isolation. In summary, the magnitude of the coefficients in both the main effects and interaction models demonstrates that last semester CGPA is the key determinant of final CGPA.

3.3. Non-academic models (main and interaction effects)

In the non-academic main effects model, non-academic predictors explain about 29% of the variance in scaled CGPA. The coefficient for Gender_Male is approximately –0.33, and that for Gaming is about –0.61, indicating that male gender and excessive gaming are associated with lower CGPA. In contrast, Computer proficiency, English proficiency, and Extracurricular participation have positive effects, with coefficients around 0.10, 0.15, and 0.47, respectively, while Income and Hometown have negligible impacts. These findings support H2, which states that socioeconomic factors like income and hometown do not significantly affect CGPA for undergraduates, as suggested by[5].

When interaction terms are introduced, the model's explanatory power increases modestly to about 32%. In this interaction model, the negative effect of Gaming becomes even more pronounced ($\approx$ –0.98), and notably, the interaction between Gaming and Extracurricular participation yields a significant positive effect (≈ 0.56). This result clearly supports H3—that while excessive gaming negatively impacts CGPA, participation in extracurricular activities can mitigate this effect [6, 7]. Additionally, the effect of Gender_Male becomes slightly stronger in the interaction model ($\approx$ –0.37), further underlining its significance.

3.4. Interactive Model (academic and non academic factors and interactions)

The results for the iterative models are shown in Figure 2. In the baseline model (Model Y_1), Last CGPA ($\beta_2 \approx 0.894$) is the most significant predictor of scaled CGPA, while Preparation ($\beta_1 \approx 0.068$) also has a meaningful positive effect. This supports Hypothesis H1, which hypothesizes that previous academic performance is a key indicator of final academic performance. In Model Y_2, we introduce an interaction between Preparation and Gender_Male ($\beta_8 \approx 0.026$), which shows a small positive moderation, indicating that the benefits of preparation may be slightly stronger for male students. Model Y_3 adds the Preparation × Computer interaction ($\beta_9 \approx 0.008$), which has a minimal effect, suggesting computer proficiency does not significantly alter the benefit of preparation. In Model Y_4, the Preparation × English interaction ($\beta_{10} \approx$ –0.015) shows a very small negative moderation, implying that preparation may be slightly less effective for students with stronger English proficiency. Model Y_5 incorporates the Last × Gaming interaction ($\beta_{11} \approx 0.032$), indicating that students with high previous CGPA may mitigate any negative impact from gaming. Model Y_6 introduces the Last × Gender_Male interaction ($\beta_{12} \approx 0.075$), suggesting that male students derive a somewhat greater benefit from prior CGPA compared to female students. Model Y_7 adds Last × Extra_Yes ($\beta_{13} \approx -0.017$), a small negative interaction, implying that extracurricular involvement slightly dampens the positive influence of prior CGPA. In Model Y_8, Last × English ($\beta_{14} \approx 0.008$) adds only a marginal effect, and finally, in Model Y_9, Last × Computer ($\beta_{15} \approx$ –0.009) is also marginal.

Overall, the R^2 remains stable across Models Y_5 to Y_9 (≈ 0.861 or 86.1%), indicating that although these additional interactions offer a slightly better fit, their contributions are modest. The consistently high coefficient of the Last CGPA, along with the positive effect of Preparation, affirms H1. Meanwhile, the interaction effects, especially those involving gender, support H4 by showing that academic benefits, like Last semester's marks and preparation, are moderated by non-academic factors.

Model comparisons in Figure 3, show that the Academic Interaction Model provides the best fit ($AIC = 430.146, BIC = 470.020, R^2 \approx 0.861$ or 86.1%), with Model Y_6 performing nearly as well. In contrast, models with no academic content perform poorly ($R^2 \approx 0.285$–0.320) (28.5% to 32%), with a much higher AIC / BIC and a lower predictive power. To summarize:

- Best Model: The Academic Interaction Model, based on its consistently strong fit indices and near-top R^2.
- Worst Models: The Non-Academic models, which explain little variance and have poor fit metrics.

4. DISCUSSION

Our findings broadly reinforce the longstanding hypothesis that academic performance, particularly last semester's CGPA, is the dominant driver of overall student success. Consistent with prior literature, students' past academic achievements and preparation efforts stand out as key predictors, overshadowing most non-academic influences. In our analysis, academic performance also appears to act as a confounder by masking the independent effects of non-academic factors when included in the model.

At the same time, the non-academic factors, although weaker overall, reveal some meaningful patterns. A novel hypothesis emerging from our work is that extracurricular involvement moderates the negative impact of excessive gaming, suggesting that balanced extracurricular participation may help mitigate detrimental leisure behaviors. Additionally, our examination of interactions indicates that preparation and attendance can be even more effective when aligned with a student's prior academic standing. These findings confirm the central role of academic performance while introducing novel evidence of context-dependent effects from non-academic factors. It highlights the value of targeted support services that integrate both academic and lifestyle dimensions of student engagement. Specifically, our results advocate for a personalized approach involving early academic monitoring, such as tracking first-year GPA [8] and timely interventions to support at-risk students. Access to historical student CGPA data may inadvertently lower performance [9]; thus, institutions should manage such data propagation in a way that it strengthens academic confidence rather than discouraging students. Personalized academic advising and organized course selection guidance are examples of tailored interventions that enable them to make well-informed decisions. In the end, these results open the door to data-driven, well-planned interventions that maximize student achievement.

5. REFERENCES

[1] Hasan Tahmid, Hasan Md. Mahmud, and Manzoor Tahbib, "Student Performance Metrics Dataset — data.mendeley.com," `https://data.mendeley.com/datasets/5b82ytz489/1`, 2024.

[2] Rens van de Schoot, David Kaplan, Jaap Denissen, Jens B Asendorpf, Franz J Neyer, and Marcel A G van Aken, "A gentle introduction to bayesian analysis: applications to developmental research," *Child Dev.*, vol. 85, no. 3, pp. 842–860, May 2014.

[3] Stan Dev Team, "Cmdstanpy 1.2.5," *https://mc-stan.org/cmdstanpy/*, 2024.

[4] Geraint Johnes and Kwok Tong Soo, "Grades across universities over time," *The Manchester School*, vol. 85, no. 1, pp. 106–131, 2017.

[5] Selcuk R. Sirin, "Socioeconomic status and academic achievement: A meta-analytic review of research," *Review of Educational Research*, vol. 75, no. 3, pp. 417–453, 2005.

[6] Claudia Buchmann, Thomas A. DiPrete, and Anne McDaniel, "Gender inequalities in education," *Annual Review of Sociology*, vol. 34, no. Volume 34, 2008, pp. 319–337, 2008.

[7] Douglas A. Gentile, Hyekyung Choo, Albert Liau, Timothy Sim, Dongdong Li, Daniel Fung, and Angeline Khoo, "Pathological video game use among youths: a two-year longitudinal study," *Pediatrics*, vol. 127, no. 2, pp. e319–329, Feb. 2011.

[8] Abdulhakeem Nurudeen, Abdulnaser Fakhrou, Nuruddeen Lawal, and Sara Ghareeb, "Academic performance of engineering students; a predictive validity study of first-year gpa and final-year cgpa," Mar 2023.

[9] Sorathan Chaturapruek, Thomas S Dee, Ramesh Johari, René F Kizilcec, and Mitchell L Stevens, "How a data-driven course planning tool affects college students' GPA," in *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, New York, NY, USA, June 2018, ACM.