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 employ a Bayesian statistical framework implemented in Stan to answer our primary research question: What are the key determinants of a student's academic performance in higher education as indicated by CGPA? Further, we seek to answer several secondary questions: Do academic and non-academic factors exert similar or distinct influences on student performance? Are prior academic records, such as secondary school grades or the previous semester's CGPA, reliable predictors of final academic outcomes? Additionally, how do interactions among non-academic factors such as extracurricular activities and gaming habits affect CGPA, and can these interactions mitigate negative influences? We also analyse socioeconomic factors such as family income, place of residence, etc., to identify trends that may influence student performance.

Using a dataset [1] from an undergraduate student survey at the University of Malaya, we analyze demographic, socioeconomic, and behavioral attributes to assess their combined effects on CGPA. Building on this framework, we investigate subgroups and interaction effects that have received limited attention in the literature to address our research questions. Our Bayesian approach offers insights that can inform 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.

Fig. 1. Overview of key predictors of final CGPA. The top left plot shows the student count per department indicating the high number of CSE students, while the remaining plots show the relationships between final CGPA and select academic and non-academic factors. The CGPA vs. Last plot (top right) reveals a strong positive trend.

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, 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. We have performed an exploratory data analysis to understand the distribution of key features against the target variable, which is shown in Figure 1.

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 ($1 : True, 0 : False$). Figure 1 shows that 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 Department as a predictor with such disparity may introduce bias and limit the model's applicability. Section 2.2 provides further details on this.

2.2. Full-featured Model

Our project starts with building a Bayesian multiple linear regression model using all the available predictors. This initial model showed some $\hat{R}$ values in the range of 3, suggesting convergence issues. We suspected this issue was likely due to the imbalance in the Department column across different groups. To test this hypothesis, we built another model excluding the Department column. We developed another model that included the department variable by grouping students into those studying CSE and those studying other subjects, using binary encoding. After comparing these two models, both produced the same R^2 values of 0.86 and demonstrated convergence with $\hat{R}$s in the range of 1. Since this feature did not significantly improve the regression fit, we dropped this predictor entirely from our subsequent study.

The Bayesian model is specified such that it assumes each observed CGPA is normally distributed around a linear combination of predictors. Specifically, the model takes the form:

$$Y_i \sim \mathcal{N}(\alpha + \mathbf{X}_i \cdot \boldsymbol{\beta}, \sigma), \tag{1}$$

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 to regularize the model and allow the data to drive inference, assigning the intercept, each coefficient, and the error term a $normal(0, 10)$ prior distribution. This approach aligns well with the Bayesian principle of integrating prior knowledge with observed data, ensuring robust and interpretable results [2].

Implementation of the model was carried out in Stan via its Python interface cmdstanpy [3]. A data dictionary was constructed, including 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 $N, K, X,$ and Y; the parameters using α, β, σ; the priors and likelihood as mentioned above. It was then compiled and sampled via the No-U-Turn Sampler Hamiltonian Monte Carlo (NUTS-HMC) method using 4 chains, each with 1000 warm-up and 2000 sampling iterations, ensuring reproducibility through a fixed random seed of 42. Convergence diagnostics, including $\hat{R}$ values and effective sample sizes, were examined to confirm the reliability of the sampling process. Following sampling, posterior samples of means, standard deviations, and confidence intervals (CI) for all parameters were drawn. Predicted CGPA values were computed using the posterior means of the intercept and coefficients, and the overall fit of the model was evaluated by comparing these predictions to the actual scaled CGPA through the calculation of an R^2 value. The model's performance was visualized by plotting predicted versus actual CGPA values. Detailed results, including the significance of individual predictors and their interactions, are discussed in Section 3.

2.3. Model - Academic Features

The impact of academic features on student performance was measured by developing two Bayesian regression models, following the same approach as the full-featured model. These models work with academic variables only and use same priors as defined for the full model in section 2.2. The first model incorporates the main effects while second one introduces interaction terms to better understand both individual and combined impacts on CGPA. The first model has 6 predictors: HSC, SSC, Preparation, Attendance, Semester, and Last. The interactive model builds on the first model by incorporating 3 more interactive features: Last $\times$ Preparation, Preparation $\times$ Attendance, and Last $\times$ Attendance.

2.4. Model - Non-academic Factors

We also developed two regression models for assessing non-academic factors. The first model considers non-academic features such as gender, income, computer access, gaming habits, employment status, English proficiency, extracurricular involvement, and hometown background. The preprocessed features were used to construct a design matrix with $K = 9$ predictors. The second model incorporated interaction terms to capture the combined effects of multiple non-academic factors. Interaction terms such as Gender_Male $\times$ Gaming, Income $\times$ Extra_Yes, and Job_Yes $\times$ Gaming were introduced to explore how these factors jointly influence CGPA.

2.5. Models - Interactive Academic and Non-academic Factors

After performing regression with all factors and evaluating academic and non-academic factors individually and interactively, we built a model that investigated the interaction effects between these two categories. We focused on the significant factors identified from the academic and non-academic models. Rather than constructing a single, overly complex model with all possible interactions, we adopted an iterative modeling approach. This approach started with a model that included one interaction term between an academic and a non-academic factor and added interaction terms progressively in subsequent models to assess their impact. For example:

- Model X_2 = Preparation + Last + Gender_Male + Computer + English + Gaming + Extra_Yes + Preparation $\times$ Gender_Male
- Model X_3 = Preparation + Last + Gender_Male + Computer + English + Gaming + Extra_Yes + Preparation $\times$ Gender_Male + Preparation $\times$ Computer.

These terms were incorporated iteratively to refine the model and allow for a focused evaluation of interaction effects. 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 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.
- 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.

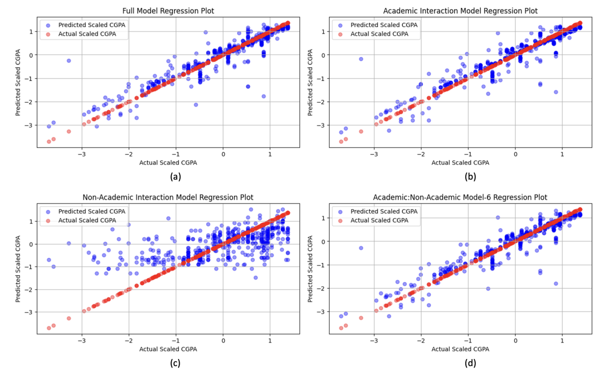

Fig. 2. Regression model performance comparison for predicted and actual scaled CGPA across different models. (a) The full-featured model demonstrates a strong alignment between the CGPAs. (b) The academic interaction model shows a similar performance. (c) The non-academic interaction model shows increased variance in predictions. (d) The academic-non-academic interactive model 6 shows a strong positive trend as well.

3.1. Full Model

Our results highlighted the dominant role of academic factors in predicting scaled CGPA. The coefficient for Last (past CGPA) is strongly positive, indicating it is the most significant predictor of scaled CGPA. This is consistent with what we have found in the literature as well [4]. Preparation also showed a moderate, positive effect. In contrast, non-academic variables exhibit CIs that crossed zero, suggesting their individual effects are statistically non-significant in the presence of strong academic predictors. The model had $\hat{R} \approx 1.00267$, which indicates good convergence, and $R^2 \approx 0.86$ (Figure 3(a)) explained approximately 86% of the variance, leaving room for further exploration of subtler non-academic effects. The regression plot in Figure 2(a) also shows the model's predictive power.

3.2. Academic models (main and interaction effects)

The academic model explains 86.07% (Figure 3(a)) of the variance in scaled CGPA. Again, the most significant predictor is Last ($\beta \approx 0.89$), while Preparation ($\beta \approx 0.07$) and Attendance ($\beta \approx 0.03$) add a modest positive influence. The effects of HSC, SSC, and Semester are minimal, with their coefficients being close to 0. Introducing interactions among Preparation, Attendance, and Last slightly improves performance with $R^2 \approx 86.3\%$, making the effect of Last even

	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	-240.15
R^2	0.861	0.861	0.863	0.285	0.320	0.861

(a)

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 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

(b)

Fig. 3. Comprehensive result table. a) Evaluation of model performance metrics (AIC, BIC, LOO, and R^2) for Full, Academic, Non-Academic, and Interaction models in comparison to other model specifications. (b) R^2 values and parameter estimates $(\beta_1 - \beta_{15})$ and intercept (α) for Models 1–9, with standard errors enclosed in parenthesis. The findings show that the model fit varies and that the coefficients are stable across specifications.

more pronounced. Meanwhile, Preparation and Attendance show negative effects when interactions with Last are considered. This implies that preparation and attendance levels in the current semester moderate the benefit of having a high GPA from the previous semester. Figure 2(b) shows a regression curve that resembles the full-featured model, further confirming the influence of the academic factors. The academic models confirmed that the past year's academic performance is the critical factor in determining final CGPA, even when non-academic factors are introduced.

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

In the absence of academic factors, non-academic predictors explain only 29% ($R^2 = 0.29$, shown in Figure 3(a)) of the variance in scaled CGPA, considerably lower than the academic model. Despite the low R^2, examining non-academic factors remains valuable for identifying potential intervention points. The non-interactive model reveals that Gender_Male and Gaming have significant negative impacts (-0.334 and -0.613), respectively, which is consistent with findings in [5] and [6]. Conversely, Computer, English, and Extracurricular showed significant positive effects $(0.101, 0.155, \text{ and } 0.468)$, respectively. However, part-time job, Income, and Hometown had no significant effects. According to [7], socioeconomic factors like income and hometown are influential for younger students, but for undergraduates, these factors have a lesser impact as students may gain independence. Adding interaction terms among these significant non-academic factors improved the model's R^2 to about 32%, as can be seen in Figure 3(a); however, the plot in Figure 2(c) shows that it still underperformed. The negative effect of Gaming persists, but notably, the interaction between Gaming and extracurricular involvement becomes significant and positive, suggesting that participation in extracurricular activities can moderate the detrimental impact of gaming, an interesting new area that can be explored. Other interactions remain non-significant.

3.4. Interactive Model (academic and non academic factors

Sequential models summarized in the Figure 3((b)) showed a modest increase in R^2 from 0.859 in the baseline Model 1 to around 0.861 by Model 6. Model 6, which included Preparation $\times$ Gender_Male, Preparation $\times$ Computer, Preparation $\times$ English, Last $\times$ Gaming, and Last $\times$ Gender_Male interactions terms, stabilized coefficient estimates with low standard errors. Models 7–9 incorporated additional interaction terms, but they yielded negligible improvements in R^2 while increasing model complexity. The effect of the Last CGPA remained robust ($\beta \approx 0.749$), and Preparation showed a slight positive effect ($\beta \approx 0.063$). Although the effect of Gaming is relatively small ($\beta \approx -0.021$), its interaction with Last ($\beta \approx 0.032$) suggests that the negative impact of gaming is moderated by academic performance. The interaction between Preparation and Gender_Male ($\beta \approx -0.014$) indicates that the benefit of preparation may be less pronounced for male students. Modest contributions arose from interactions between Computer and English. Thus, model 6 balances explanatory power and parsimony, making it the best representation of academic and non-academic interactions on scaled CGPA, as seen in Figure 2(d). Figure 3(a) clearly shows that academic models performed far better than non-academic-only models, having lower AIC/BIC values, higher LOO scores, and significantly higher R^2. The Academic Interaction Model has the lowest AIC and BIC (430.146 and 470.020), indicating the best trade-off between fit and complexity. The Full Model and interactive Model 6 have very similar performance (AIC/BIC near 455–460, $R^2 \approx 0.861$). Meanwhile, the non-academic models have much higher AIC/BIC, lower LOO scores, and low R^2 (around 0.285–0.320), making them the weakest performers overall. To summarise:

- 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] 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.

[6] 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.

[7] 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.

[8] Abdulhakeem Nurudeen, Abdulnaser Fakhrou, Nurudeen 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.