Student Mental Health and Academic Performance

Malynda Holt, Alex Osborne
College of Engineering
University of Missouri
Columbia, MO 65211
{mahnkc, aocdv} @umsystem.edu`

Abstract

We investigate the feasibility of predicting student academic performance (GPA) using mental health and lifestyle factors, with the goal of identifying key predictors and enabling early intervention strategies. Using a dataset of 760 students comprising 20 features, we engineered composite indicators such as a Stress Index, Wellness Score, and Academic-Social Imbalance to capture latent patterns in the data. Our baseline models—including Linear Regression, Random Forest, and Gradient Boosting—exhibited poor generalization, with test R^2 values below zero and high mean absolute errors. While overall predictive accuracy remained limited, the study reveals meaningful correlations between mental health metrics and GPA, offering insights for developing targeted educational support frameworks.

1 Introduction

Academic performance is influenced by psychological, social, and environmental factors. Research continuously emphasizes the role of mental health and lifestyle in shaping student outcomes, especially with the rising stress and burnout rates recorded in academia. While prior studies have examined individual correlations between well-being and GPA, relatively few have explored predictive modeling approaches to quantify and interpret these relationships systematically.

In this work, we address the problem of predicting student GPA based on a diverse set of self-reported mental health, lifestyle, and demographic features. Our motivation is twofold: first, to identify which factors are most strongly associated with academic outcomes, and second, to explore whether such data can enable meaningful early interventions for at-risk students.

We utilize a publicly available dataset of 760 students that includes over 20 features, ranging from stress levels and sleep habits to social support and study time. We employ a data-centric approach, involving extensive exploratory data analysis, feature engineering, and model evaluation using both linear and non-linear methods. To enhance interpretability and performance, we introduce composite variables that synthesize multiple signals into interpretable scores reflecting stress, wellness, and study-life balance.
Despite the complexity of the data and the limitations in modeling performance, our findings

contribute actionable insights. We demonstrate that engineered features capturing holistic wellness and balance outperform raw input features in predictive utility. Furthermore, we reflect on the inherent difficulty of modeling human academic outcomes, where latent variables, noise, and non-linear effects limit model generalizability.

This work emphasizes the importance of thoughtful feature design and simplicity in modeling complex behavioral outcomes and lays a foundation for future work focused on classification and personalized support systems in educational settings.

2 Data and Exploratory Analysis

We utilize the *Student Mental Stress and Coping Mechanisms* dataset, sourced from Kaggle, which contains self-reported survey responses from 760 students. The dataset comprises 20 features spanning mental health indicators (e.g., stress levels, anxiety, depression), lifestyle factors (e.g., sleep duration, study hours, physical activity), and demographic attributes (e.g., age, gender, family type). The target variable is academic performance, represented as a GPA on a 0–4 scale. author of the paper.

2.1 Data Overview

The dataset includes a balanced distribution of GPA values, allowing for generalization across high- and low-performing students. Most features are categorical or ordinal in nature, with a few continuous variables such as sleep hours and study time. Some columns contained free-text responses or highly sparse categories, which were removed during preprocessing to improve modeling consistency.

To ensure data quality, we conducted standard preprocessing steps, including handling missing values, encoding categorical variables, and filtering out non-numeric or ambiguous fields. Categorical variables like gender and counseling attendance were one-hot encoded, while redundant or collinear variables were dropped to reduce noise.

2.2 Exploratory Analysis

Exploratory Data Analysis (EDA) revealed several statistically significant trends. Students who reported higher sleep duration, more study hours, and stronger family support tended to have higher GPAs. Sleep and study time, in particular, showed modest but consistent linear correlations. Elevated levels of mental stress, excessive social media usage, and financial pressure were all associated with lower academic performance. Among these, mental stress showed the strongest negative correlation with GPA. Notably, students with a balanced routine outperformed peers with extreme study hours but poor health or social support.

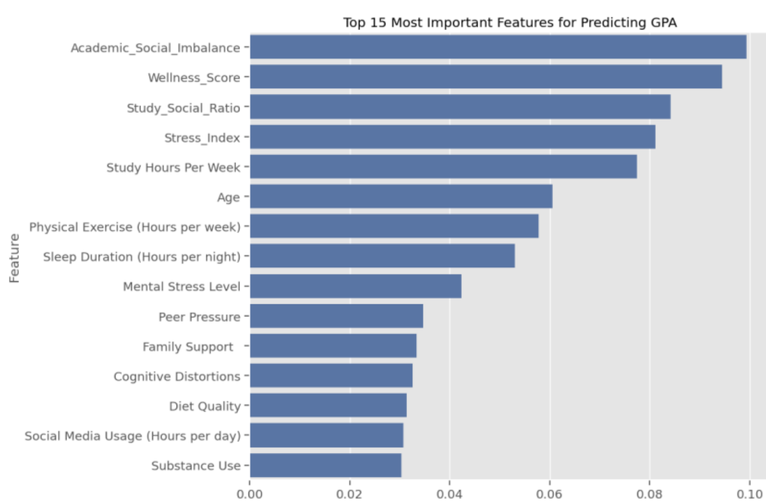

Figure 1: Feature Importance Graph

These findings motivated the creation of composite variables to better capture latent constructs such as wellness and stress, which are not fully represented by any single feature. Figure 1 (not shown here) illustrates the correlation matrix highlighting the relationships between GPA and other key variables.

In summary, the EDA validated the relevance of mental health and lifestyle features in predicting academic performance and provided a foundation for feature engineering and model design described in the next section.

3 Methodology

Our modeling pipeline was designed to extract meaningful patterns from a small, complex dataset by emphasizing interpretability, feature quality, and generalization. This section describes the key components of our approach: preprocessing, feature engineering, model selection, evaluation metrics, and refinement strategies.

3.1 Data Preprocessing

Prior to modeling, we applied several preprocessing steps to ensure data consistency. First, we removed records with missing target values and dropped records with excessive null entries. Since the dataset contained various different types of data, we classified each feature and dropped some features that could not be easily quantified. Binary features such as gender and counseling attendance were encoded one-hot. Multi-class variables with inconsistent or low representation were removed. Open ended text responses were excluded, as they could not be reliably integrated in with the data and introduced high variance. Continuous variables such as study hours, sleep duration, and stress ratings were standardized to zero mean and unit variance for models sensitive to feature scale.

3.2 Feature Engineering

Motivated by insights from our EDA, we constructed several high-level composite features to capture latent patterns not easily represented by raw variables:

- Stress Index: Aggregates responses related to academic stress, time pressure, and anxiety into a single numerical score.
- Wellness Score: Combines positive lifestyle factors (e.g., sleep hours, physical activity) and subtracts stress-related indicators to reflect overall well-being.
- Academic-Social Imbalance: Measures the deviation from a balanced lifestyle, quantifying how much academic effort dominates at the expense of leisure, social interaction, or rest.

These engineered features were intended to reflect the holistic conditions that influence academic performance, going beyond superficial metrics like hours studied. In total, we retained a reduced feature set of both raw and engineered variables for modeling.

3.3 Data Modeling

We implemented and compared several supervised learning algorithms for regression:

- Linear Regression: Used as a baseline for interpretability and simplicity.
- Random Forest Regressor: A non-linear ensemble method capable of modeling interactions and non-monotonic relationships.
- Gradient Boosting Regressor: Designed to capture complex patterns through iterative learning.

To evaluate generalization and robustness, we later introduced:

- Ridge and Lasso Regression: Regularized linear models that penalize complexity and help mitigate overfitting, especially valuable in small datasets.

3.4 Evaluation Metrics

Model performance was assessed using three standard regression metrics:

- Mean Absolute Error (MAE): Reflects the average absolute deviation from the true GPA, offering an intuitive measure of error magnitude.
- Root Mean Squared Error (RMSE): Penalizes larger errors more than MAE, emphasizing consistency.
- R^2 Score: Measures the proportion of variance explained by the model; negative values indicate performance worse than predicting the mean.

These metrics were evaluated on both training and test sets to track overfitting and generalization.

3.5 Refinement

To improve model performance and reduce overfitting, we implemented several refinements:

- Cross-Validation: We adopted 5-fold cross-validation to obtain more stable performance estimates and reduce variance due to data splitting.
- Feature Selection: Recursive feature elimination and correlation analysis were used to identify a subset of high-signal variables.
- Hyperparameter Tuning: Grid search was applied to key hyperparameters such as regularization strength (α), tree depth, and learning rate. Simpler models with stronger regularization and shallower trees consistently outperformed more complex variants.

Together, these steps were designed to balance predictive power with interpretability and robustness, given the challenges posed by the data.

4 Results and Discussion

4.1 Baseline Model Performance

We first evaluated three baseline models: Linear Regression, Random Forest, and Gradient Boosting. Across all models, test performance was notably poor. Linear Regression achieved a test R^2 of -0.073, while Random Forest and Gradient Boosting fared even worse with R^2 scores of -0.121 and -0.169, respectively. In practical terms, all models had a mean absolute error (MAE) between 1.2 and 1.3, indicating an average prediction error of approximately one full letter grade on a 0–4 GPA scale.

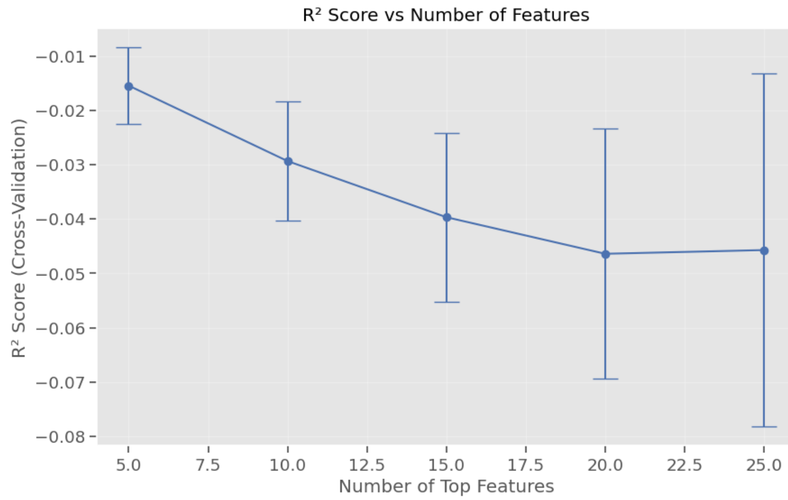

Figure 2: R Score vs Number of Features

These results highlight the inherent difficulty of the task. Not only were the models unable to outperform a naive mean prediction, but they also demonstrated signs of overfitting—particularly Random Forest, which achieved a training R^2 of 0.847. This gap between training and test performance suggests that the models were too complex relative to the dataset size and signal-to-noise ratio.

4.2 Feature Importance and Interpretability

Despite low predictive accuracy, feature importance analysis revealed consistent patterns across models. Our engineered features, Wellness Score, Academic-Social Imbalance, and Stress Index, accounted for approximately 35% of total predictive importance, confirming the value of composite metrics that capture holistic aspects of student well-being. Raw features such as study hours, physical exercise, and age were also moderately predictive, though they appeared more sensitive to model variance.

The consistent emergence of engineered features in top importance rankings validates our feature engineering strategy and suggests that GPA outcomes are more accurately described by high-level constructs than by isolated behaviors.

4.3 Refinement Outcomes

To improve model robustness, we introduced regularized linear models (Ridge and Lasso Regression) and implemented 5-fold cross-validation. These refinements led to modest but meaningful improvements. The best-performing model was Ridge Regression with strong regularization ($\alpha = 100$), which achieved a test MAE of 1.16 and R^2 of -0.028. This represented a significant reduction in error compared to baseline models.

An unexpected but important finding was that models trained on a reduced set of five high-signal features consistently outperformed those using the full feature set. This aligns with the classic bias-variance tradeoff and reinforces the value of careful feature selection when working with small, noisy datasets.

4.4 Challenges in Modeling GPA

The difficulty in achieving high predictive accuracy is consistent with prior findings in educational and psychological research. Many important variables, such as instructor quality, coursework difficulty, motivation, and home environment, were not available in the dataset. Additionally, academic performance is influenced by complex, non-linear, and often stochastic factors that are not easily captured in survey-based features.

Furthermore, in social science domains, R^2 values as low as 0.1–0.3 are often considered meaningful. From this perspective, even a modest reduction in error and identification of key contributing factors represents valuable progress, particularly when the goal is to inform interventions rather than to make precise predictions.

4.5 Implications and Insights

Although the predictive power of our models was limited, the project produced several actionable insights. The consistent importance of holistic wellness and lifestyle balance points to promising directions for student support services. For example, early screening tools based on our engineered features could be used to identify students at risk of underperformance, not by predicting exact GPA, but by flagging imbalances or stressors.

Additionally, our results suggest that future models could benefit from shifting toward classification (e.g., low-, medium-, high-risk GPA tiers) rather than regression, potentially improving both interpretability and utility in practical settings.

References

[1] Salah Uddin Ahmed Shuvo. (2023) *Student Mental Stress and Coping Mechanisms*. https://www.kaggle.com/datasets/salahuddinahmedshuvo/student-mental-stress-and-coping-mechanisms. Accessed April 24, 2025.