

Machine Learning-Based Prediction of Student GPA from Academic Behaviors

Group 6

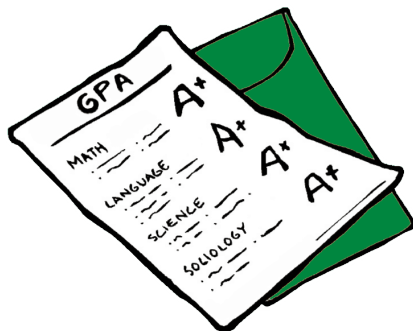
Vương Hồng Minh Nguyễn Quang Anh Đỗ Nguyễn gia Như
Trần Thanh Phát Nguyễn Quang Minh

April 2025

Table of content

- 1 Problem definition
- 2 Dataset
- 3 Methodology
- 4 Result
- 5 Model Evaluation
- 6 Demo

Problem definition



Project Goals

- **Predict** student academic performance
- **Analyze** learning patterns
- **Identify** at-risk students
- **Provide** early interventions

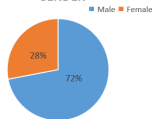
Summary Dashboard



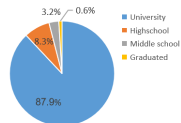
NUMBER OF SURVEYS

157

GENDER

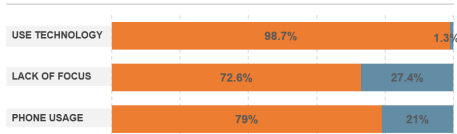


CURRENT EDUCATION LEVEL

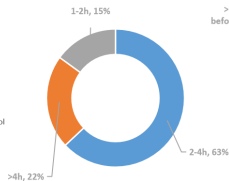


THE IMPACT OF EXTERNAL FACTORS

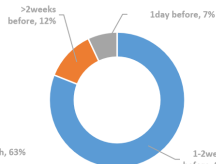
● TRUE ● FALSE



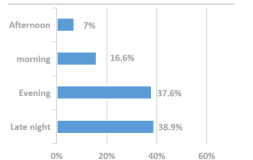
DAILY STUDY DURATION



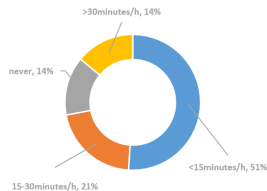
EXAM PREPARATION APPROACH



MOST EFFECTIVE STUDY TIME

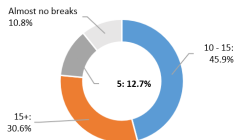


SOCIAL MEDIA USAGE DURING STUDY TIME

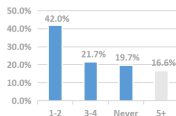



Summary Dashboard

 Study break frequency (min/h)



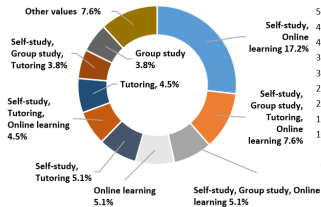
 Weekly exercise frequency(times)



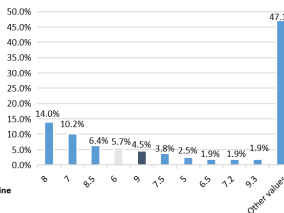
 Daily sleep duration (hours)



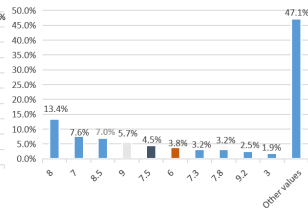
Preferred study method

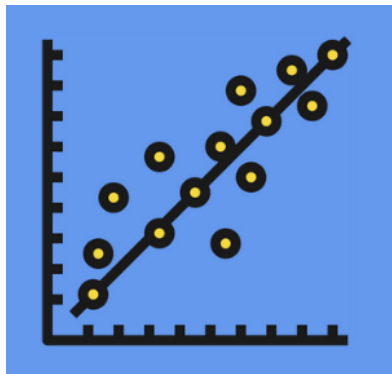


Latest semester GPA



Latest academic year GPA





Linear Regression



with some libraries

① Data preprocessing

- The dataset is divided into three types: **numerical**, **ordinal**, and **nominal** features.
- Missing values are handled using `SimpleImputer`.
- The data is split into 80% training and 20% testing sets.

② Training the model using Linear Regression

- Apply a bounded Linear Regression model with pipeline integration.

③ Evaluating the model

- Use MAE, MSE, and R^2 to assess performance.

④ Predicting output using the test set

- Generate predicted GPA values and compare with actual outcomes.

● Predicted & Actual GPA:

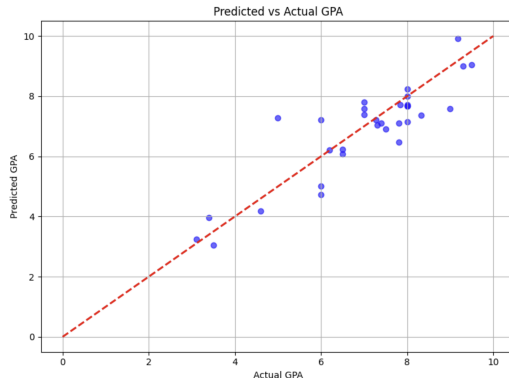
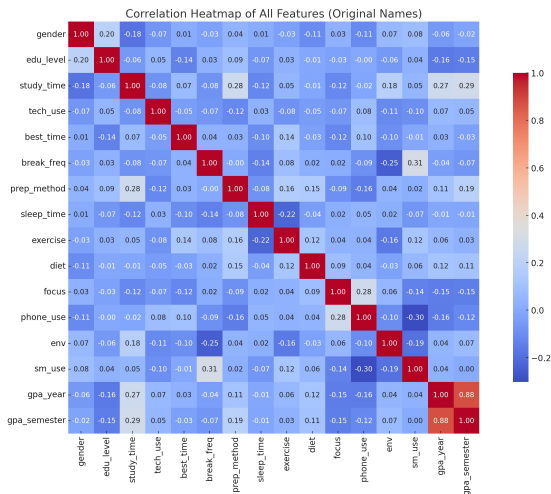


Figure 3.1: Predicted vs Actual GPA

Correlation Heatmap of all Features:



Evaluate Model Results

- **Mean Absolute Error (MAE)** ≈ 0.597 — The average of absolute differences between predicted and actual GPA values.
- **Mean Squared Error (MSE)** ≈ 0.5972 — Gives more weight to large errors, useful for identifying outliers.
- **Coefficient of Determination (R^2 score)** ≈ 0.755 — Indicates that the model explains about 77.5% of the variance in GPA.

⇒ **The model shows good predictive performance overall.**

Disadvantages

- Only works well if data is simple and linear
- Can be wrong if data has outliers
- Cannot learn complex patterns
- Gets confused if features are too similar

Future Work

- Try better models like Random Forest or Neural Network
- Add more useful features (class time, mental health, etc.)



Thank you
for listening