CampusPulse Initiative – Project Report
Student Relationship Prediction from Lifestyle & Academic Data
Indian Institute of Technology, Guwahati

Author:
Srimaya Mohapatra
244104010

Table of Contents

Section Details

1. Introduction

- Brief on the CampusPulse initiative
- Objective: Understand factors influencing relationships using data and machine learning
- Your goal: Build a pipeline from data cleaning to explainable predictions

2. Dataset Overview

- Mention source (anonymized IITG survey)
- Number of records (e.g., 649 students)
- Column types: demographics, academics, lifestyle, etc.

- Explain Feature_1 to Feature_3 were anonymized

3. Task 1: EDA & Modeling

3.1 Level 1 – Variable Identification

- Techniques used: correlation heatmaps, scatter plots
- Findings: Feature_1 likely → alcohol use, Feature_2 → screen time, Feature_3 → stress
- Justification using G1–G3 and behavior variables

3.2 Level 2 – Data Cleaning

- Missing columns: Fedu, famsize, etc.
- Imputation strategies: mean for numeric, mode for categorical
- Justification for each

3.3 Level 3 – Exploratory Questions

Example questions:

- Does going out affect grades?
- Do students with more absences perform worse?
- Does parental education influence G3?
- Each supported with plots and interpretation

3.4 Level 4 – Modeling

- Algorithms used: Logistic Regression, Decision Tree, Random Forest, Naive Bayes
- Accuracy summary table
- Best model: Random Forest (~67.92%)
- Confusion matrix + class balance notes

3.5 Level 5 – SHAP Explainability

- Global importance: goout, freetime, Feature_3 ranked highest
- Local force plots: one student predicted yes, one no
- Insight: lifestyle > academics in predicting relationships

3.6 Bonus – Boundary Match

- Plots guessed: KNN, Decision Tree, RF, SVM, Logistic Regression
- Reasoning: shape of boundary (linear vs nonlinear, blocky vs smooth)

4. (If Task 2 was assigned, summarize it here similarly.)
5. Summary of Experiments

- Tried multiple models
- Handled missing values
- Visual storytelling (EDA)
- Used SHAP for human-centered interpretation

6. Conclusion

- Key takeaway: Social/lifestyle features are most predictive
- Value: Transparency using SHAP helps understand student well-being
- Future work: Add sentiment data, expand features

7. Appendix (optional)

- Include key plots: heatmap, SHAP summary, model comparison table