CampusPulse Initiative – Task 1 Report

# 1. Introduction

This report outlines the analysis and modeling process for Task 1 of the Machine Learning challenge during Coding Week 2025 at IIT Guwahati. The CampusPulse Initiative focuses on understanding student life using real, anonymized survey data and involves tasks such as data cleaning, exploratory data analysis (EDA), predictive modeling, and model interpretation.

# 2. Level 1: Variable Identification Protocol

In this phase, we performed exploratory data analysis to identify the anonymized features: Feature\_1, Feature\_2, and Feature\_3. Using correlation analysis, scatter plots, and domain intuition, we proposed likely interpretations for each feature based on statistical behavior and associations with known variables.

# 3. Level 2: Data Integrity Audit

We conducted a thorough audit of missing values and data inconsistencies. Features with missing values were identified, and appropriate imputation strategies were applied based on the nature of the data (e.g., mean for continuous features, mode for categorical ones). Justifications for each strategy were documented within the code comments.

# 4. Level 3: Exploratory Insight Report

This stage involved asking and answering key questions about student behavior and wellness using visualizations. We explored variables like academic workload, screen time, stress levels, and sleep habits. Insights were presented through bar charts, violin plots, and interpreted in context.

# 5. Level 4: Relationship Prediction Model

The core machine learning task involved building classification models to predict whether a student is in a romantic relationship based on their academic and lifestyle patterns. We applied models such as Logistic Regression, Decision Trees, and Random Forest, and evaluated them using accuracy, F1 score, and confusion matrices.

# 6. Level 5: Model Reasoning & Interpretation

Transparency in modeling was achieved using SHAP for interpretability. Global feature importances and local explanations were generated for individual predictions. Decision boundaries were plotted for selected model-feature pairs to visualize how classifiers differentiate between classes.

# 7. Conclusion

The CampusPulse Initiative successfully demonstrated data cleaning, analysis, and modeling capabilities on real-world student data. Insights from the EDA, along with predictive modeling and interpretation, provide a solid foundation for building decision-support tools aimed at improving student wellness.