Machine Learning CW 2025 – Project Report

By: Smai Debbarma  
B.Tech Chemical Engineering, IIT Guwahati  
Coding Club CW 2025 Submission

# Task 1: CampusPulse – Predicting Relationships from Student Life Data

## Level 1: Variable Identification Protocol

We performed exploratory data analysis (EDA) to identify anonymized features (Feature\_1, Feature\_2, Feature\_3). Using correlation matrices, histograms, and scatter plots, we inferred possible meanings for each feature:  
- Feature\_1 likely represents sleep duration due to high correlation with energy level and CGPA.  
- Feature\_2 seems to relate to screen time or phone usage.  
- Feature\_3 could reflect social activity frequency based on distribution trends.

## Level 2: Data Integrity Audit

We identified missing values in a few categorical and numerical columns. Imputation strategies included:  
- Mean imputation for continuous variables (e.g., sleep hours).  
- Mode imputation for categorical features (e.g., study time).  
- Dropped rows only where too many nulls existed.

## Level 3: Exploratory Insight Report

We explored the dataset through 5 key questions:  
1. Does screen time affect CGPA?  
2. Is stress level different for students in relationships?  
3. Is sleep duration linked to academic performance?  
4. Are students with more social events more likely in a relationship?  
5. Does weekday alcohol use (Dalc) influence relationship probability?  
Plots were generated using scatterplots, violin plots, and bar charts. Clear trends were observed, such as a slight inverse relation between screen time and CGPA.

## Level 4: Relationship Prediction Model

We used classification models to predict whether a student is in a relationship. Models tried:  
- Logistic Regression  
- Random Forest  
- Support Vector Machine (SVM)  
  
Best model: Random Forest with ~83% accuracy on test data.  
Features like screen time, social frequency, and alcohol use showed significant influence.

## Level 5: Model Interpretation

We used SHAP values to understand model decisions:  
- Global feature importance was plotted showing social activity and alcohol use among top factors.  
- Local SHAP analysis explained predictions for two students (1 Yes, 1 No).  
Decision boundaries were visualized for SVM and Logistic models using 2D projections.

# Task 2: The Rise of the WeatherMind – AI Agent System

## Level 1: Core Activation

Implemented a chatbot agent in LangGraph using an LLM backend and a calculator tool to handle math queries. The graph was rendered to visualize tool integration and flow.

## Level 2: Senses of the World

Added two external tools:  
- Weather extractor using an API.  
- Fashion trend recommender using real-time location parsing.  
The agent could now respond to location-based or context-rich prompts.

## Level 3: Judgement and Memory

Routing logic enabled tool selection based on user intent. A memory mechanism was implemented to remember prior conversation context. Demonstrated a multi-turn conversation where user goals evolved and were remembered by the bot.

## Level 4: Multi-Agent Evolution

Created three agents with different roles:  
- Planner  
- Weather analyst  
- Trend advisor  
Each agent could trigger tools and pass responses back for coordinated replies. Prompt chaining and fallback behavior were tested.

# Conclusion

This project provided hands-on experience with both data analysis and agent-based AI systems. We explored end-to-end pipelines, interpreted models with SHAP, and designed LangGraph-based AI agents. ChatGPT was used for code generation ideas (~30–40%) while debugging, visualizations, and experimentation were manually driven.

# Appendix: Sample Outputs from Task 1 Notebook

Below are selected outputs extracted directly from the Jupyter Notebook used in Task 1:

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 649 entries, 0 to 648  
Data columns (total 33 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 school 649 non-null object   
 1 sex 649 non-null object   
 2 address 649 non-null object   
 3 famsize 599 non-null object   
 4 Pstatus 649 non-null object   
 5 Medu 649 non-null int64   
 6 Fedu 576 non-null float64  
 7 Mjob 649 non-null object   
 8 Fjob 649 non-null object   
 9 reason 649 non-null object   
 10 guardian 649 non-null object   
 11 traveltime 576 non-null float64  
 12 failures 649 non-null int64   
 13 schoolsup 649 non-null object   
 14 famsup 649 non-null object   
 15 paid 649 non-null object   
 16 activities 649 non-null object   
 17 nursery 649 non-null object   
 18 higher 573 non-null object   
 19 internet 649 non-null object   
 20 famrel 649 non-null int64   
 21 freetime 604 non-null float64  
 22 goout 649 non-null int64   
 23 Dalc 649 non-null int64   
 24 health 649 non-null int64   
 25 absences 580 non-null float64  
 26 G1 649 non-null int64   
 27 G2 614 non-null float64  
 28 G3 649 non-null int64   
 29 Feature\_1 611 non-null float64  
 30 Feature\_2 603 non-null float64  
 31 Feature\_3 610 non-null float64