## Instructions:

Save all files using the format: **RollNumber_Name** *(*Example: ***19MCMI01_ABCDEF_Lab2****)*

- Submit the following files:
  - Executable code file (.py or .ipynb)
  - Document in LaTeX (.pdf)
    **Note:** All plots, analysis, results and discussions should be included in the document.
- Send the files to ***scis2026mllab@gmail.com*** ***by Friday (23rd Jan 2026 11:59 PM)***
- Ensure that the code executes without errors and the document is properly formatted before submission.

## Task:

## Analyse how student habits influence their academic performance

1. Download the Student Habits vs Academic Performance dataset from Kaggle.
2. Load the dataset.
3. Analyze and print the information and statistical summary of the dataset.
4. **Perform preprocessing to handle missing, duplicate, and null values (if any), and convert the exam score to labels based on the SCIS M.Tech grade scale.**
5. Perform exploratory data analysis using pair plots.
6. Split the dataset into an 80–20 train–test split.
7. **Apply decision tree classification model.**
8. **Plot the confusion matrix (understand the details) and evaluate the model using Accuracy, Precision, Recall, and F1-score metrics for the classification task.**
9. Repeat the experiment with different train–test split ratios:
   - 90–10
   - 70–30
   - 60–40
   - 50–50
10. Plot and analyze how the evaluation metrics (**Accuracy, Precision, Recall, and F1-score metrics**) vary across different train–test splits (x-axis: train–test split, y-axis: metrics).

**Libraries to be used:** **NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn.**