Machine Learning Project:

Group members: Mahnoor Arif (bsdsf22a010)**, Sidra Siddique** (bsdsf22a030)

Project Title

Predicting Student Academic Performance Using Machine Learning

1. Problem Statement

In educational systems, early identification of students who are at risk of underperforming is critical for timely intervention. Traditional methods of student evaluation are reactive rather than predictive, lacking the foresight to proactively support struggling students. With the increasing availability of educational data, there is an opportunity to apply machine learning techniques to predict academic performance and uncover hidden patterns that influence learning outcomes.

2. Objectives

- To develop a machine learning model that can accurately predict student academic performance based on demographic, social, and academic features.
- To analyze which factors most significantly influence student outcomes.
- To assist educational institutions in identifying at-risk students early for targeted intervention.
- To compare multiple machine learning algorithms for best performance.

3. Proposed Methodology

Step 1: Data Collection

- *Dataset: **Student Performance Dataset** from UCI Machine Learning Repository*
  - *URL: https://archive.ics.uci.edu/dataset/320/student+performance*
  - *The dataset includes attributes such as gender, age, study time, failures, family support, absences, and final grades (G1, G2, G3).*

Step 2: Data Preprocessing

- Handle missing values (if any)
- Convert categorical variables using label encoding/one-hot encoding
- Normalize/scale features
- Split data into training and testing sets (e.g., 80/20 split)

Step 3: Model Selection and Training

- Models to be evaluated:
  - **Linear Regression** (for baseline)
  - **Decision Tree Regressor**
  - **Random Forest Regressor**

    - **Support Vector Machine (SVM)**
- Evaluation metrics:
    - Mean Absolute Error (MAE)
    - Root Mean Squared Error (RMSE)
    - R^2 Score

Step 4: Feature Importance & Analysis

- Use models like Random Forest or SHAP values to analyze feature impact
- Visualize insights using heatmaps and importance plots

Step 5: Model Deployment

- If time allows, build a simple UI to input student attributes and predict outcomes using Flask or Streamlit

5. Dataset Description

- Total Records: 649 students
- Features: 33 attributes per student
- Output variable: **G3** (Final Grade)
- Dataset contains two subsets: Math and Portuguese language performance
- Types of attributes:
    - Demographic (e.g., age, gender)
    - Academic (e.g., previous grades, absences)
    - Social (e.g., parental education, support, relationship status)

6. Expected Outcomes

- A trained machine learning model that predicts student final grades with high accuracy
- Identification of key predictors of academic success or failure
- Comparative analysis of different algorithms' performance
- (Optional) A lightweight predictive tool for academic advisors

7. Project Timeline

Weeks	Task Summary
Week 1	Literature review, dataset acquisition, preprocessing
Week 2	Model selection, training, and evaluation
Week 3	Performance comparison, feature importance analysis
Week 4	finalization, (optional) UI deployment, presentation