Project Proposal

# Project Title:

Student Score Analyzer & Grade Predictor

# Project Members:

Batool Binte Fazal (2024140)

Rida Syed (2024540)

# 1. Project Objectives

This project aims to develop an interactive web application that allows users to:  
- Upload student performance datasets.  
- Visually explore academic trends and score distributions.  
- Predict student scores in a selected subject using a regression model.  
- Classify students as pass or fail using a logistic regression classifier.  
- Input hypothetical student data and receive real-time predictions.  
- View performance-based study suggestions.  
  
The app leverages Streamlit to build a dynamic dashboard that combines machine learning and user-friendly design in the education domain.

# 2. Methodology

## Dataset

Source: Students Performance in Exams – Kaggle <https://www.kaggle.com/datasets/spscientist/students-performance-in-exams>  
Includes approximately 1000 records with demographic data and scores in math, reading, and writing.

## Tools & Libraries Used

- pandas – for data loading, cleaning, and feature engineering.  
- numpy – for numeric operations and derived metrics.  
- matplotlib – for generating bar charts and visual summaries.  
- scikit-learn – for training linear and logistic regression models.  
- Streamlit – for building an interactive and responsive web interface.

## Implementation Steps

1. **Data Upload & Cleaning**  
   - Users upload a CSV file.  
   - The app removes null values and duplicates.  
   - Encodes categorical variables using LabelEncoder.  
     
   **2. Feature Engineering**  
   - Calculates average score across math, reading, and writing.  
   - Optionally computes:  
    • study\_efficiency = average score / (study time + 1)  
    • study\_engagement = study time × (average score / 100)  
    • education\_factor = average score / (parental education level + 1)  
    **3. Data Visualization**  
   - Bar charts for average scores by gender (if present).  
   - Correlation matrix displayed using pandas styling.  
     
   4**. Machine Learning Models**  
   - Linear Regression is used to predict a selected subject score.  
    • RMSE (Root Mean Square Error) is reported.  
   - Logistic Regression classifies students as pass/fail (≥50).  
    • Accuracy is calculated and shown.  
    **5. User Input & Prediction**  
   - A form allows users to input new student data.  
   - The app predicts both the subject score and pass/fail outcome.  
   - Based on the score, the app gives one of three performance suggestions:  
    • < 50: Suggests time management and tutoring.  
    • 50–75: Encourages more practice.  
    • > 75: Praises strong performance.

**6. Grade Assignment Functionality**:

A new feature was added to classify students into grades based on their predicted scores:

* A for scores >= 90
* B for scores >= 80
* C for scores >= 70
* D for scores >= 60
* F for scores < 60

# 3. Expected Outcomes

- A complete Streamlit dashboard that allows:  
 • Data upload and preview after cleaning.  
 • Visualization of key academic metrics.  
 • Score prediction using regression.  
 • Pass/fail prediction using classification.  
 • Real-time prediction for custom input with actionable feedback.

# 4. Key Learning Outcomes

- Data preprocessing and exploratory analysis using pandas.  
- Building and evaluating ML models with scikit-learn.  
- Feature engineering for education-related metrics.  
- Designing responsive and functional Streamlit web apps.

**Bonus features included:**

* Study tips for underperforming students
* Score insights by gender
* Grade prediction with letter grades

This project demonstrates foundational technical and user interface development skills, while maintaining an accessible level of complexity suitable for early-stage academic projects and professional growth opportunities.