# **Predicting Student Exam Scores Using Regression Models**

## **Objective**

The goal of this project is to predict students' final exam scores using input features such as hours studied, attendance, and previous exam performance. By applying regression techniques, this model aims to assist educators in identifying key performance factors and supporting students at academic risk.

## **Dataset Used**

* **Source:** UCI Student Performance Dataset
* **File Used:** student-mat.csv (Math performance)
* **Total Records:** 395 students
* **Features Used:**
  + studytime (approximate hours studied weekly)
  + failures (number of past class failures)
  + absences (total number of absences)
  + G1, G2 (grades from first and second evaluation)
* **Target Variable:** G3 (final grade)
* Visualized relationships using **pairplots** between studytime, G1, G2, and G3
* Used a **correlation heatmap** to identify strong predictors of final grade
  + Strongest correlations observed:
    - G2 ➝ G3: 0.8
    - G1 ➝ G3: 0.7
  + Negative correlation between failures and G3

### **1. Linear Regression**

* **R² Score:** 0.78
* **MSE:** 4.46
* **Top Coefficient:** G2

**2. Random Forest Regressor**

* **R² Score:** 0.87
* **MSE:** 2.61
* **Top Feature Importance:**
  + G2: 81%
  + Absences: 12%
  + G1: 3%

Random Forest performed significantly better in capturing nonlinear relationships.

A scatter plot was generated to compare **actual vs predicted G3 scores** using the Random Forest model. Most predictions clustered close to the actual values line, indicating good accuracy.

* **Previous performance (G2 and G1)** are highly indicative of final exam scores.
* **High number of absences or past failures** slightly lower the predicted outcome.
* Surprisingly, **studytime had minimal predictive power**, possibly due to it being self-reported and categorical.