# Student Performance Prediction Based on Study Habits

# Objective

# The aim of this project is to predict whether a student will pass or fail based on multiple academic, demographic, and lifestyle-related factors. This prediction can help educators and institutions identify students who may need extra support.

# Dataset Used

# The dataset contains a variety of features that influence student performance. Key attributes include:

# Gender

# Age

# Study Time

# Failures

# Family Support

# Internet Access

# Absences

# Health

# Parental Education Level

# School Type

# Travel Time

# Address Type (Urban/Rural)

# Final Grade

# Target Variable: Pass/Fail (binary classification)

# Model Chosen

For this project, we used:

1. **Logistic Regression** – As a baseline for binary classification.
2. RandomForestClassifier - To improve accuracy and handle non-linearity in feature space.

* **Methodology**

Exploratory Data Analysis (EDA)

Data Preprocessing

Model Training

Model Evaluation

# Performance Metrics

1. **Accuracy**: The proportion of correctly predicted outcomes out of all predictions.
2. **Precision**: The proportion of true positives among all predicted positives.
3. **Recall**: The proportion of true positives identified out of all actual positives.
4. **F1-Score**: The harmonic means of precision and recall for balanced performance.
5. **Confusion Matrix**: A table showing true/false positives and negatives to evaluate model performance.

* **Model Evaluation Results:**

1. For Logistic regression
   * 1. Accuracy: 0.2916666666666667
     2. Precision: 0.2569444444444444
     3. Recall: 0.2916666666666667
2. For Random Forest
   * 1. Accuracy: 0.375
     2. Precision: 0.2886904761904762
     3. Recall: 0.375

* **Challenges**
* Handling class imbalance.
* Selecting meaningful features that contribute to academic outcomes.
* Balancing model complexity vs interpretability.
* **Learnings:** 
  + Gained practical experience in **data preprocessing**, including encoding, scaling, and handling categorical variables.
  + Understood how different **features impact student performance** and the importance of EDA in revealing hidden insights.
  + Learned how to apply and compare **Logistic Regression** and **RFC** for binary classification.
  + Discovered how **feature scaling** significantly affects RFC performance.
  + Enhanced skills in evaluating model performance using **accuracy, precision, recall, F1-score, and confusion matrix**.
  + Developed a user-friendly **Streamlit app** to deploy the model and make real-time predictions.

# Conclusion

This project demonstrates the effective use of Logistic Regression to predict student performance (Pass/Fail) using a variety of academic and lifestyle features. The model achieved high accuracy and reliable evaluation metrics, indicating strong predictive capability. The process covered essential steps in a machine learning pipeline—from data preprocessing to model training and evaluation—culminating in a simple and interactive Streamlit app for real-time predictions. This solution can help identify at-risk students early and support data-driven decisions in the education domain.