Student Performance Prediction Project

Problem Statement

We aim to predict whether a student will pass the math exam (score >= 40) using demographic and preparation data. This is a binary classification problem, where the model outputs 1 for pass and 0 for fail.

Dataset

Dataset: 'StudentsPerformance.csv' from Kaggle

Features used: gender, race/ethnicity, parental level of education, lunch, test preparation course

Target: pass_math (1 = pass, 0 = fail)

Tools & Libraries

Python, pandas, seaborn, matplotlib, scikit-learn, RandomForestClassifier

Project Steps

1. Loaded and explored the dataset.

2. Visualized relationships using Seaborn (boxplots, countplots).

3. Encoded categorical variables.

4. Engineered the target column (pass_math).

5. Trained a Random Forest Classifier.

6. Evaluated model accuracy and performance.

Results

The model achieved an accuracy of around 85% (varies slightly on each run). Confusion matrix and classification report show good precision and recall.

Prediction Example

Given input features: [male, group B, bachelor's degree, standard lunch, completed test preparation], the model predicts the student will PASS the math exam.

Conclusion

This project demonstrates the ability to build, evaluate, and explain a basic machine learning model using real-world education data. The insights can help schools focus on preparation programs to improve outcomes.