# 📊 Maths Grade Predictor

## 🎯 Project Objectives

- To explore and analyze student performance data in Mathematics.

- To identify key factors that influence final Mathematics grades (\*\*G3\*\*).

- To build a \*\*Linear Regression model\*\* that predicts student grades based on significant predictors.

- To implement a \*\*Random Forest Regression model\*\* and compare its performance with Linear Regression.

- To evaluate model accuracy using metrics such as \*\*Mean Absolute Error (MAE), Mean Squared Error (MSE), and R² Score\*\*.

- To derive actionable insights that can inform targeted academic interventions for improved student outcomes.

## 📂 Dataset

- \*\*File:\*\* `student-mat.csv`

- \*\*Source:\*\* UCI Machine Learning Repository

- \*\*Description:\*\* Contains demographic, academic, and social attributes of Portuguese secondary school students and their Mathematics grades (G1, G2, G3).

## 🗂️ Project Structure

Maths\_GradePredictor/

├── data/

│ └── student-mat.csv

├── notebooks/

│ └── maths\_grade\_prediction.ipynb

├── README.md

- \*\*data/\*\*: Contains the dataset CSV file

- \*\*notebooks/\*\*: Jupyter notebook for data analysis and modeling

- \*\*README.md\*\*: Project overview, results summary, and documentation

## 💡 Skills Demonstrated

- Data cleaning and preparation using \*\*Pandas\*\*

- Exploratory data analysis (EDA) and visualization

- Correlation analysis to identify key predictors

- Building and evaluating \*\*Linear Regression\*\* models

- Implementing \*\*Random Forest Regression\*\* for performance comparison

- Model evaluation using \*\*MAE, MSE, and R² metrics\*\*

- Version control using \*\*Git and GitHub\*\*

- Clear documentation and project structuring

## 🛠️ Tools Used

- Python

- Pandas

- Scikit-learn

- Matplotlib / Seaborn

- Jupyter Notebook

- Git & GitHub

## 📈 Analysis & Results

### 🔬 Correlation Analysis

Key correlations with \*\*final Mathematics grade (G3)\*\*:

- G2: 0.92

- G1: 0.83

- Study time: 0.25

- Failures: -0.39

- Age: -0.11

- Absences: -0.09

✅ \*\*Interpretation:\*\* G1 and G2 grades are strong predictors of final grade, while failures have a moderately negative correlation.

### 🤖 Linear Regression Results

- \*\*Mean Absolute Error (MAE):\*\* 0.73

- \*\*Mean Squared Error (MSE):\*\* 1.33

- \*\*R-squared (R² Score):\*\* 0.86

✔️ \*\*Interpretation:\*\* The Linear Regression model explains approximately \*\*86% of the variance\*\* in final grades, with an average error of 0.73 grade points.

### 🌲 Random Forest Regression Results

- \*\*Mean Absolute Error (MAE):\*\* 0.82

- \*\*Mean Squared Error (MSE):\*\* 1.98

- \*\*R-squared (R² Score):\*\* 0.80

✔️ \*\*Interpretation:\*\* The Random Forest model performs slightly worse than Linear Regression in this dataset context.

### ⚖️ Model Comparison Summary

| Metric | Linear Regression | Random Forest |

|--------|-------------------|---------------|

| MAE | 0.73 | 0.82 |

| MSE | 1.33 | 1.98 |

| R² | 0.86 | 0.80 |

✅ \*\*Conclusion:\*\* Linear Regression slightly outperformed Random Forest in predicting Mathematics grades for this dataset.

## 🚀 Future Enhancements

- Implement \*\*cross-validation\*\* to improve model reliability

- Explore additional algorithms (e.g. Gradient Boosting, Support Vector Regression)

- Deploy the model as an \*\*interactive web application or dashboard\*\*

- Combine with Portuguese dataset for holistic student performance insights

- Conduct feature engineering for improved predictive accuracy

## 👩‍💻 Author

Adebanke Sobanke

\*Data Science and Technical Product Management Enthusiast\*