Predicting Student Performance with Machine Learning: A Data-Driven Approach

Introduction

Student performance prediction is a crucial aspect of educational analytics, helping educators and institutions make informed decisions. In this project, I built and deployed a machine learning model using **Linear Regression** to predict student performance based on various academic and socio-economic factors. This post details the dataset, model training, key insights, and deployment process.

Project Background

The dataset, sourced from **Kaggle**, contained **17 features** related to student demographics, academic history, and behavioral factors. The primary objective was to analyze these features, identify key performance indicators, and build a predictive model that provides actionable insights for educators.
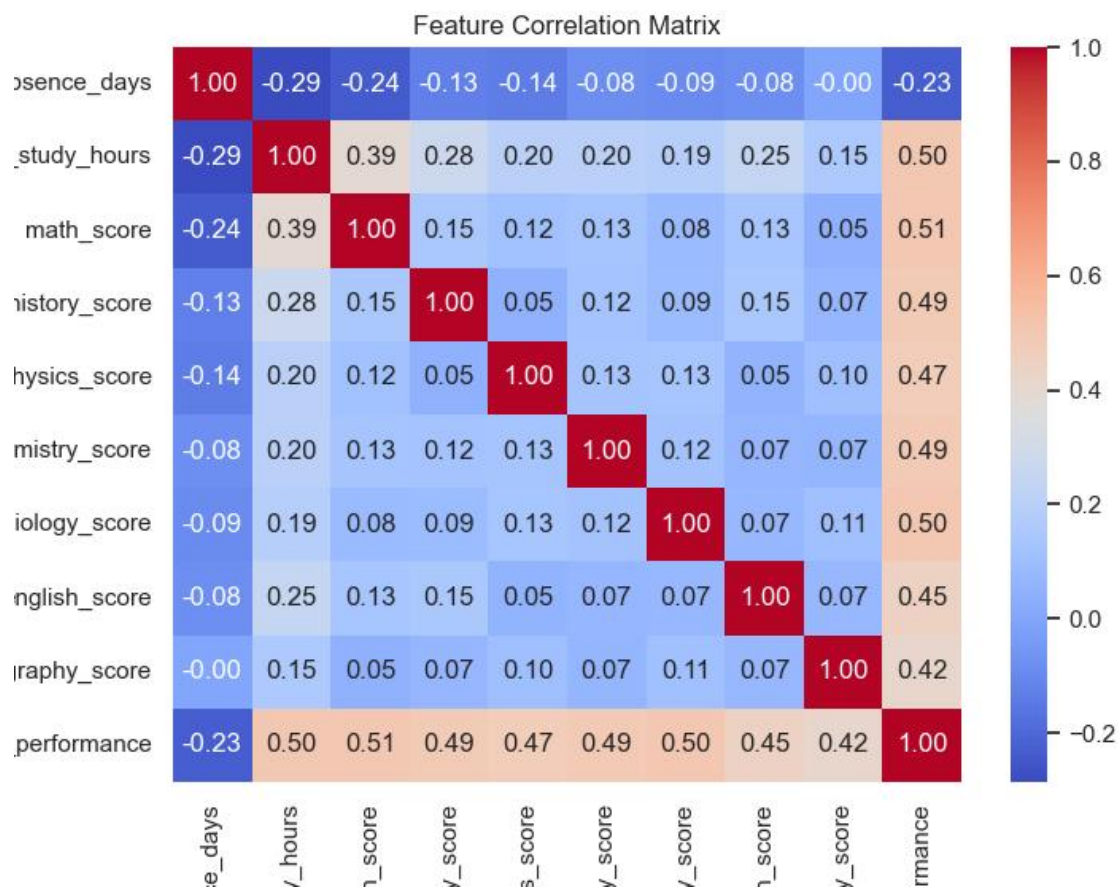
Data Collection & Preparation

✓ **Dataset Source:** Kaggle
✓ **Features:**

- **Academic Performance:** Math, History, Physics, Chemistry, Biology, English, and Geography scores.
- **Behavioral Factors:** Absence days, part-time job, weekly self-study hours, extracurricular activities.
- **Demographic Information:** Gender, career aspirations.

Preprocessing Steps:

- Dropped redundant columns (e.g., ID, names, email).
- Encoded categorical variables.
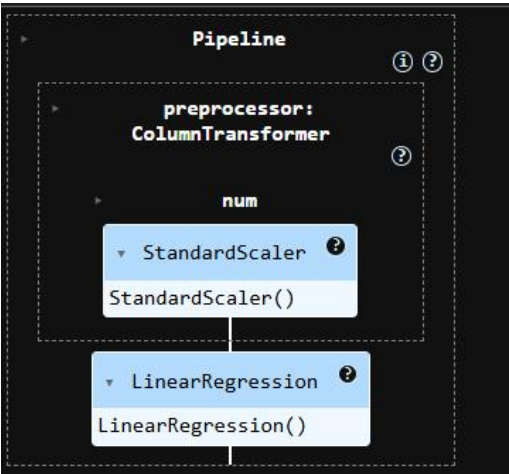- Scaled numerical features for better model performance.

Exploratory Data Analysis (EDA)

Key findings from the data analysis:

Feature Correlation Matrix

	e_days	/_hours	1_score	/_score	s_score	/_score	/_score	1_score	/_score	rmance
sence_days	1.00	-0.29	-0.24	-0.13	-0.14	-0.08	-0.09	-0.08	-0.00	-0.23
study_hours	-0.29	1.00	0.39	0.28	0.20	0.20	0.19	0.25	0.15	0.50
math_score	-0.24	0.39	1.00	0.15	0.12	0.13	0.08	0.13	0.05	0.51
history_score	-0.13	0.28	0.15	1.00	0.05	0.12	0.09	0.15	0.07	0.49
physics_score	-0.14	0.20	0.12	0.05	1.00	0.13	0.13	0.05	0.10	0.47
mistry_score	-0.08	0.20	0.13	0.12	0.13	1.00	0.12	0.07	0.07	0.49
iology_score	-0.09	0.19	0.08	0.09	0.13	0.12	1.00	0.07	0.11	0.50
nglish_score	-0.08	0.25	0.13	0.15	0.05	0.07	0.07	1.00	0.07	0.45
raphy_score	-0.00	0.15	0.05	0.07	0.10	0.07	0.11	0.07	1.00	0.42
performance	-0.23	0.50	0.51	0.49	0.47	0.49	0.50	0.45	0.42	1.00

- **Weekly Self-Study Hours (0.50)** showed a strong positive correlation with total student performance, emphasizing the importance of study habits.
- **Math Score (0.51)** and **Biology Score (0.50)** were the strongest predictors of total academic performance.
- **Absence Days (-0.23)** negatively impacted performance, indicating that frequent absences lead to lower scores.

Modeling & Analysis

```
Pipeline                           ⓘ ⓘ

  preprocessor:
  ColumnTransformer
                                   ⓘ
      num

    ▾ StandardScaler    ❷
  StandardScaler()

    ▾ LinearRegression  ❷
  LinearRegression()
```

- **Algorithm Used:** Linear Regression within a **Pipeline**
- **Pipeline Components:**

  - Data preprocessing (encoding categorical variables, feature scaling)
  - Model training and evaluation

Performance Evaluation:

- The model generalized well without underfitting or overfitting.
- Achieved a **high r² score**, indicating a strong relationship between predicted and actual student scores.
- No significant residual errors, confirming the reliability of predictions.

Deployment

- **Platform:** Deployed using **Streamlit**, making the model accessible for real-time predictions.
- **User Interface:** Simple input fields allow educators to enter student details and receive predicted performance scores instantly.

Key Insights & Future Enhancements

- The model can help educators identify at-risk students early and implement interventions.
- Future improvements could include:

  - Incorporating additional datasets to enhance predictive accuracy.
  - Experimenting with other regression techniques like Ridge or Lasso Regression.
  - Integrating real-time student feedback for continuous model improvement.

Conclusion & Call to Action

This project highlights the potential of machine learning in education. If you're interested in exploring the model, check it out on **[GitHub/Live App] (https://github.com/DGideonnene/ML-Predicting-Model-for-Student-Performance).** Feel free to share your thoughts in the comments or connect with me for more insights!