

Student Score Prediction

Using Linear Regression & Excel Dashboard

Author: Shreya Sah

GitHub: https://github.com/Shreya-Sah/guvi_project2

Project Overview

- Objective: Predict a student's Final Score using Hours Studied & Attendance Percentage.
- Approach:
 - - Data preprocessing in Python
 - - Model training with Linear Regression
 - - Evaluation (R^2 , MAE)
 - - Interactive Excel dashboard for predictions

Machine Learning Model Development

- Model Type: Linear Regression
- Inputs: Hours Studied, Attendance %
- Target: Final Score
- Steps:
 - Load dataset (student_scores.csv) in Python

1	Hours_Studied	Attendance	Final_Score
2	5	90	85
3	3	60	55
4	6	95	90
5	2	40	42
6	7	98	92
7	1	30	35
8	4	80	75
9	5	85	82
10	3	50	48
11	6	92	88
12	2.5	65	60
13	4.5	78	72

Continued...

- Clean data (remove missing values) and removed duplicates.
- **Input:**

```
# STEP 2: Load dataset
df = pd.read_csv('student_scores.csv')
print("First few rows of the dataset:")
print(df.head())

# STEP 3: Clean the data (Check and handle missing values)
print("\nMissing values in each column:")
print(df.isnull().sum())

# Remove any rows with missing values (optional: df.fillna() could be used instead)
df.dropna(inplace=True)
```

- **Output:**

```
First few rows of the dataset:
Hours_Studied  Attendance  Final_Score
0             5.0         90           85
1             3.0         60           55
2             6.0         95           90
3             2.0         40           42
4             7.0         98           92

Missing values in each column:
Hours_Studied      0
Attendance         0
Final_Score        0
dtype: int64
```

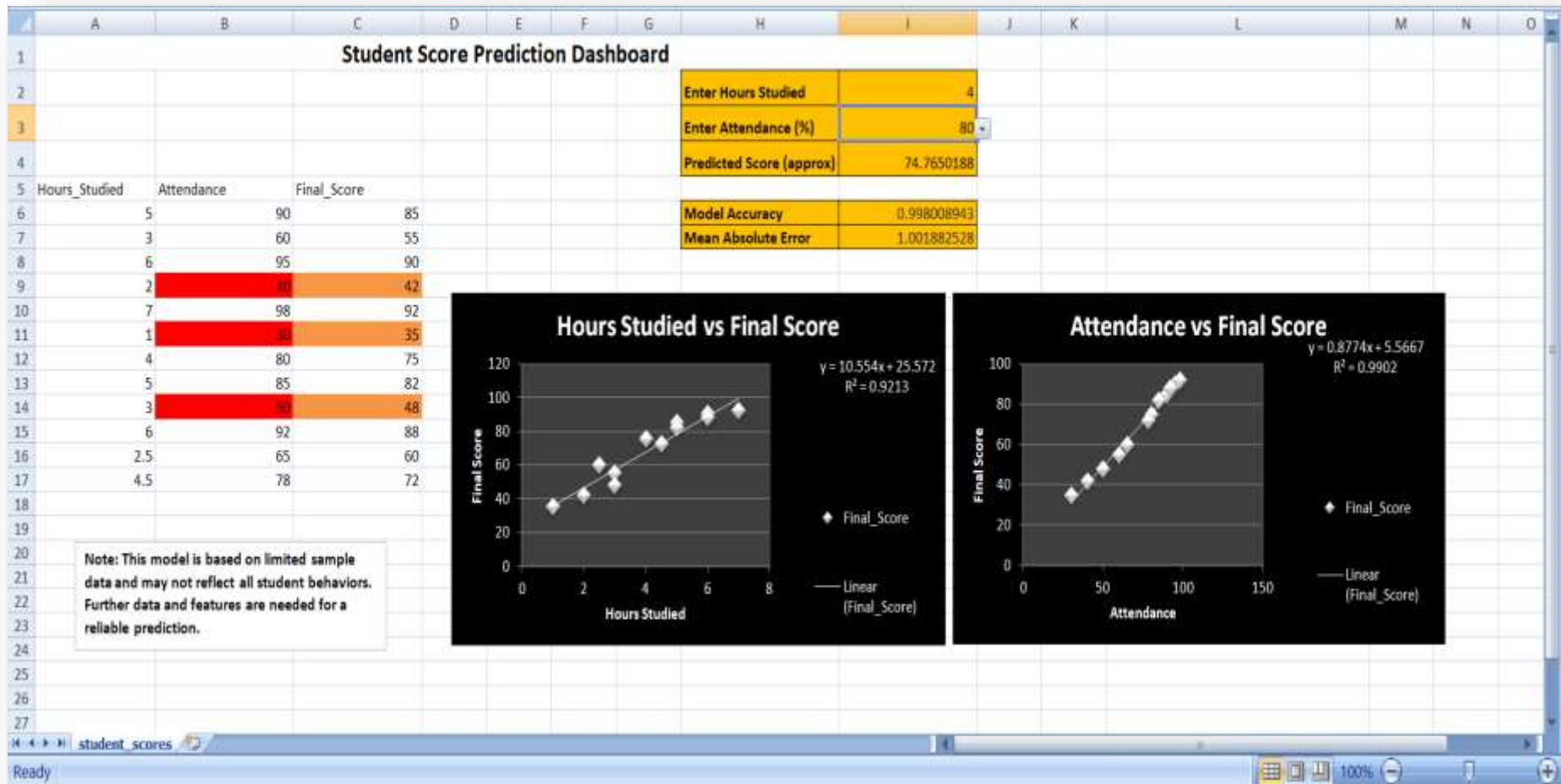
Continued...

- Train/Test split -> (80/20)
- Fit Linear Regression model
- Evaluation Metrics:
 - R^2 Score: Measures variance explained
 - MAE: Average prediction error

Dashboard Quality & Interactivity

- Interactive Excel dashboard
- Features:
 - Dropdown menus for Hours Studied & Attendance
 - Auto-updating Predicted Score (based on model)
 - Scatter plots with trendlines
 - Model accuracy metrics displayed
- Conditional Formatting in Excel Dashboard:
 - A. Low Attendance (<60%) → highlighted in red
 - B. Low Score (<50%) → highlighted in orange

Dashboard Image



Integration of SQL/Python into Data Workflow

- **Python Tasks:**
 - Data loading and cleaning (pandas)
 - Exploratory Data Analysis (seaborn, matplotlib)
 - Model training & evaluation (scikit-learn)
- **Outputs:**
 - Predicted scores for given inputs
 - R^2 Score and MAE for accuracy evaluation
 - Results linked to Excel dashboard for visualization

Data Interpretation & Insight Communication

- **Observations from Dataset:**
- Positive correlation between Hours Studied & Final Score
- Positive correlation between Attendance & Final Score
- Both features combined improve prediction accuracy
- **Communication:**
- Pairplots for visual trend detection
- Dashboard charts for quick insights
- Highlighting weak performance areas with formatting

Ethical & Bias Awareness in Model/Visualization

- **Limitations:**
- Small dataset — may not generalize
- Only two predictors (Hours, Attendance) used
- Ignores factors like motivation, study resources, health
- **Bias Prevention:**
- Clearly label predictions as estimates
- Use transparent note in dashboard
- Avoid misleading visualizations

Files in the Repository

- `guvi_project2.ipynb` – Python notebook for model training & evaluation
- `student_scores.csv` – Dataset with Hours Studied, Attendance, Final Score
- `student_dashboard.xlsx` – Interactive Excel dashboard with dropdowns, charts, and prediction formula
- `README.md` – Project documentation

Technologies Used

- Python Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn
- Excel Tools: Data Validation (dropdowns), Scatter plots + trendlines, Conditional formatting, Formula-based predictions

Workflow

- 1. Load & clean data (remove missing values)
- 2. Exploratory Data Analysis (pairplots, trends)
- 3. Train Linear Regression Model
- 4. Evaluate performance (R^2 , MAE)
- 5. Build Excel Dashboard for end-user interaction

Excel Dashboard Features

- Dropdown menus for Hours Studied & Attendance
- Automatic predictions from trained model
- Two scatter plots: Hours Studied vs Final Score, Attendance vs Final Score
- Conditional formatting to highlight low performance
- Transparent note on limitations

Model Performance

- R^2 Score: 0.9980 (indicates excellent fit)
- MAE: 1.00 (average error in score points)
- Interpretation: Predictions are highly accurate for this dataset

Model Evaluation Metrics:

R^2 Score: 0.9980089427424917

Mean Absolute Error: 1.0018825275981154

Predicted Final Score for 4 hours studied and 80% attendance: 74.66

Conclusion & Future Scope

- Achieved high accuracy in predicting scores
- Easy-to-use Excel dashboard
- Future scope: Larger datasets, more input features, integration into a web/app interface

Contact

- Author: Shreya Sah
- GitHub: https://github.com/Shreya-Sah/guvi_project2

***THANK
YOU***