ibimbing

Predicting Student Scores with Machine Learning

Presented by Rifa Sadiqa

ibimbing

Table of Contents

- Introduction
- Exploratory Data Analysis
- Splitting Data
- Regression Models Used
- Dataset Visualization
- Model Evaluation
- Conclusion & Takeaways
- Certificate of Completion
- Let's Connect!

Introduction

Why Predicting Student Scores Matters

In today's data-driven world, educational performance analytics has become a powerful tool to understand and support student success.
One of the key questions in this area is:

Can we predict a student's score based on their study hours?
This project explores that question using machine learning regression techniques.

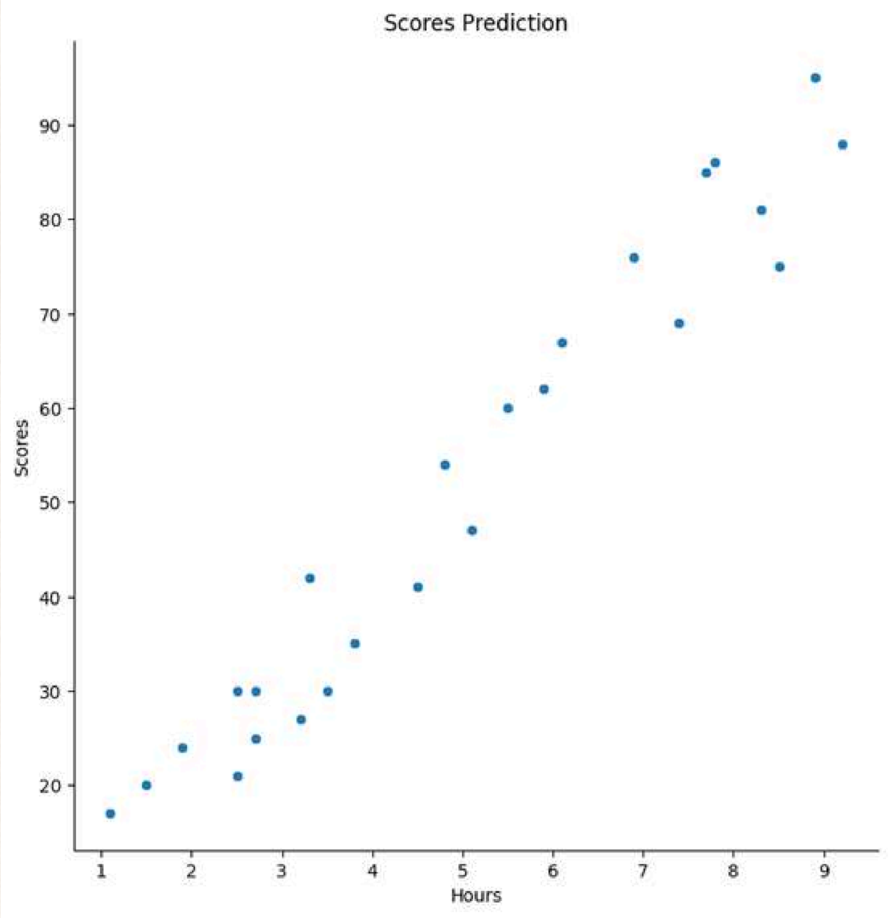
Why This Topic?

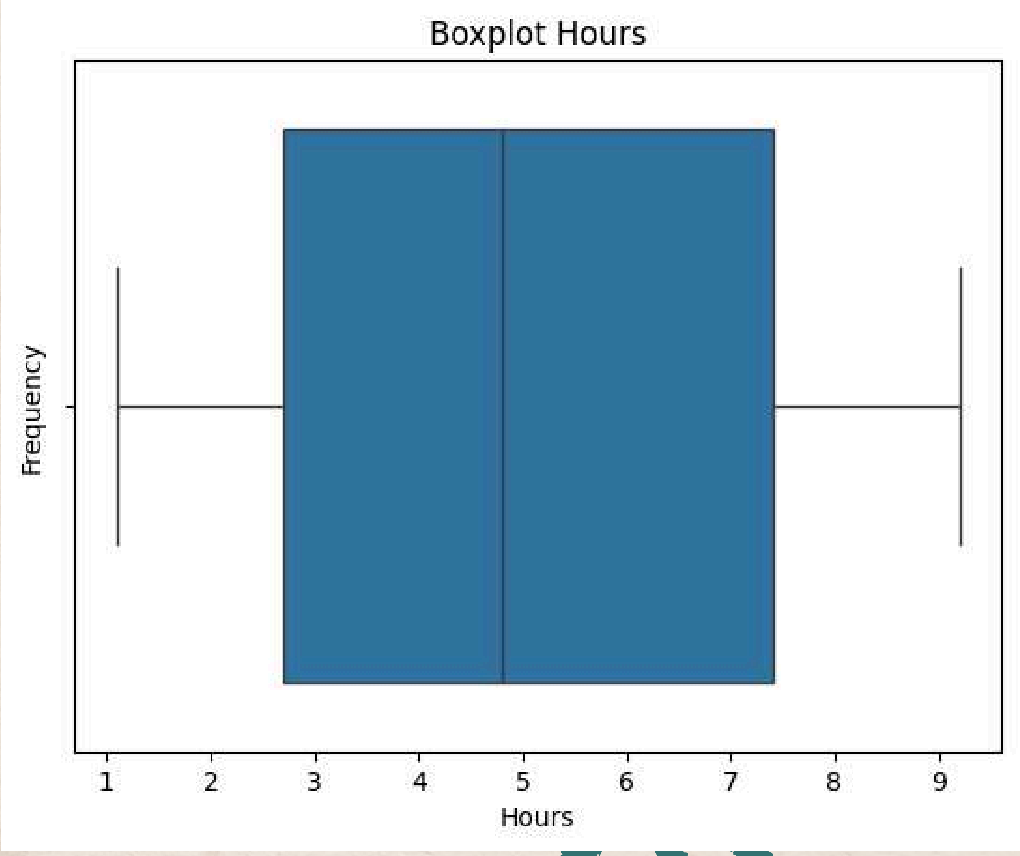
- Helps identify learning patterns and optimize study plans
- Demonstrates real-world application of supervised learning
- A simple yet meaningful case to strengthen my data science foundation

↩ Return to TOC

Exploratory Data Analysis (EDA)

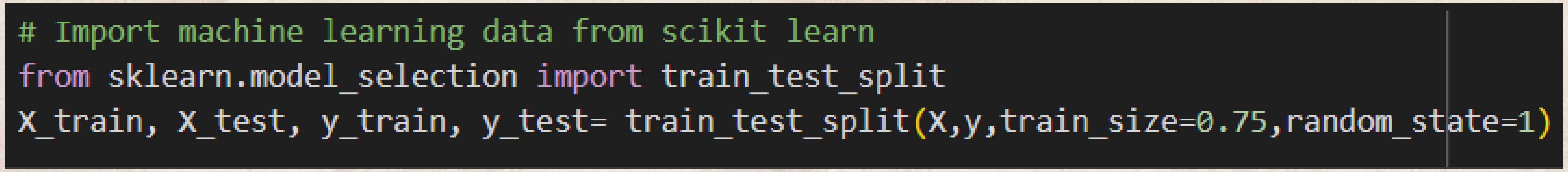
- Dataset: Contains 25 data points showing the number of study hours vs corresponding scores.
- Checked for missing values → Not found
- Checked for duplicates → All unique records
- Outlier detection using Boxplot → No extreme outliers
- Relationship between Hours and Scores visualized using scatter plot → Reveals a strong positive linear correlation





Return to TOC

Splitting Data



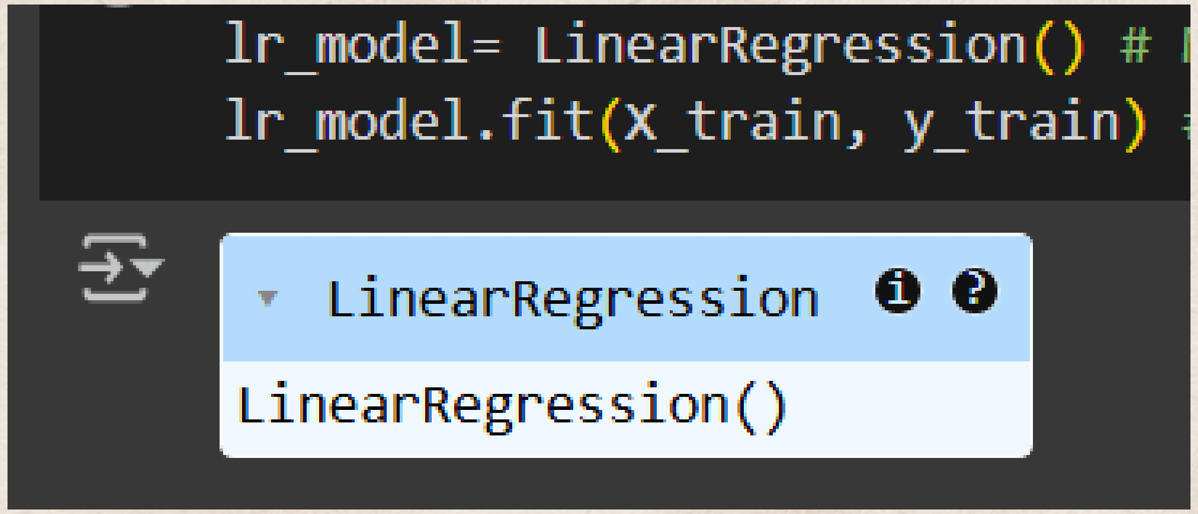
```
# Import machine learning data from scikit learn
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test= train_test_split(X,y,train_size=0.75,random_state=1)
```

To evaluate model performance objectively, I split the dataset into training and testing sets using an 75-25 split.

- Training Set: 75% of the data
- Testing Set: 25% of the data
- Applied standard data preprocessing techniques before modeling.

Return to TOC

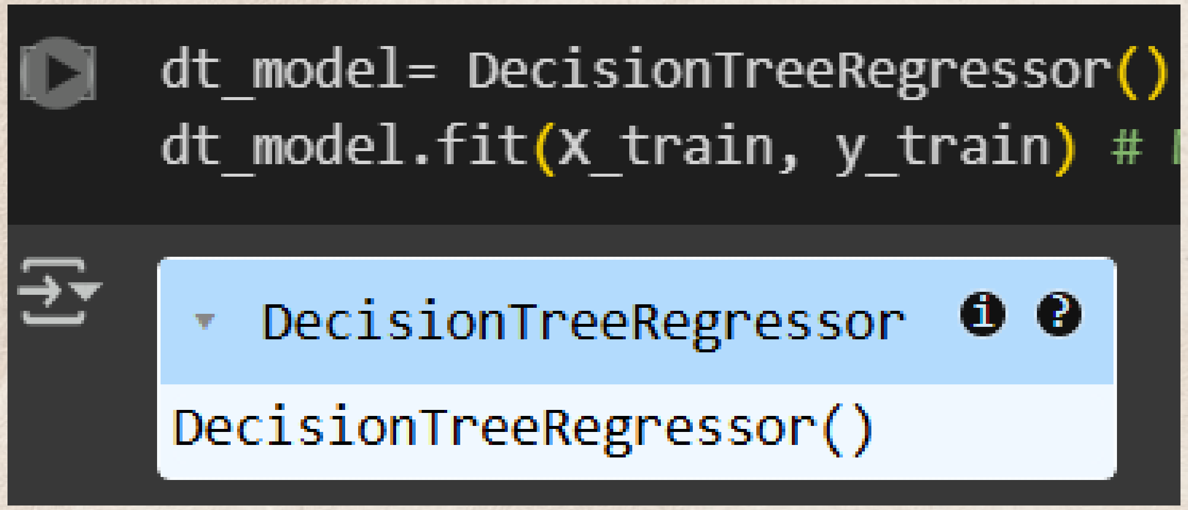
Regression Models Used



```
lr_model= LinearRegression() #
lr_model.fit(X_train, y_train)
```

```
▾ LinearRegression
LinearRegression()
```

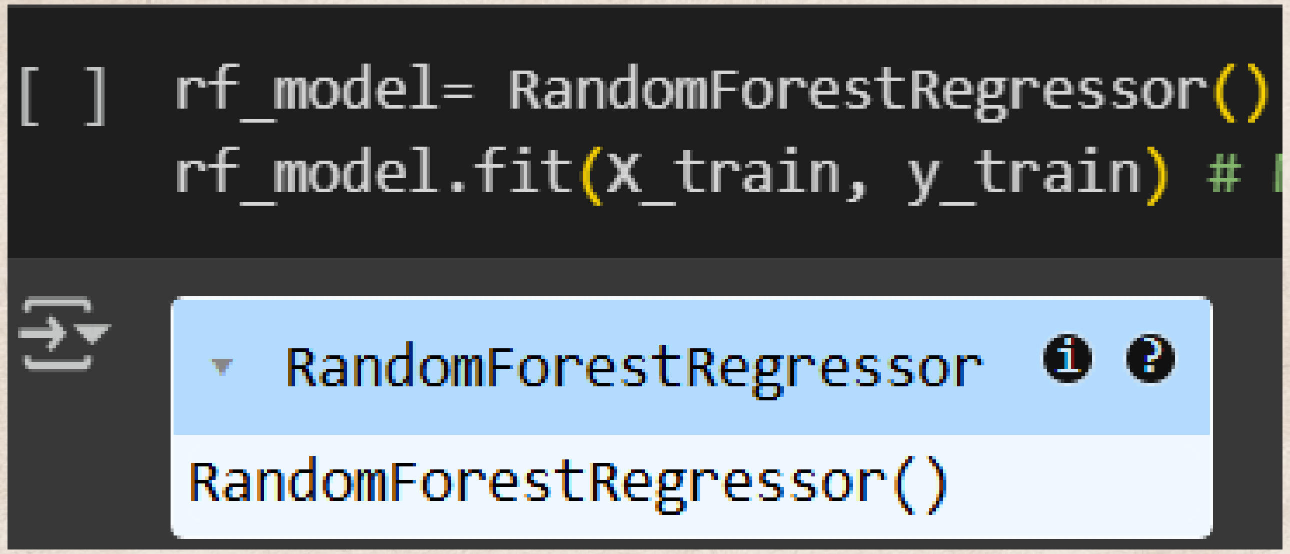
1. Linear Regression



```
dt_model= DecisionTreeRegressor()
dt_model.fit(X_train, y_train) #
```

```
▾ DecisionTreeRegressor
DecisionTreeRegressor()
```

2. Decision Tree Regressor



```
[ ]  rf_model= RandomForestRegressor()
     rf_model.fit(X_train, y_train) #
```

```
▾ RandomForestRegressor
RandomForestRegressor()
```

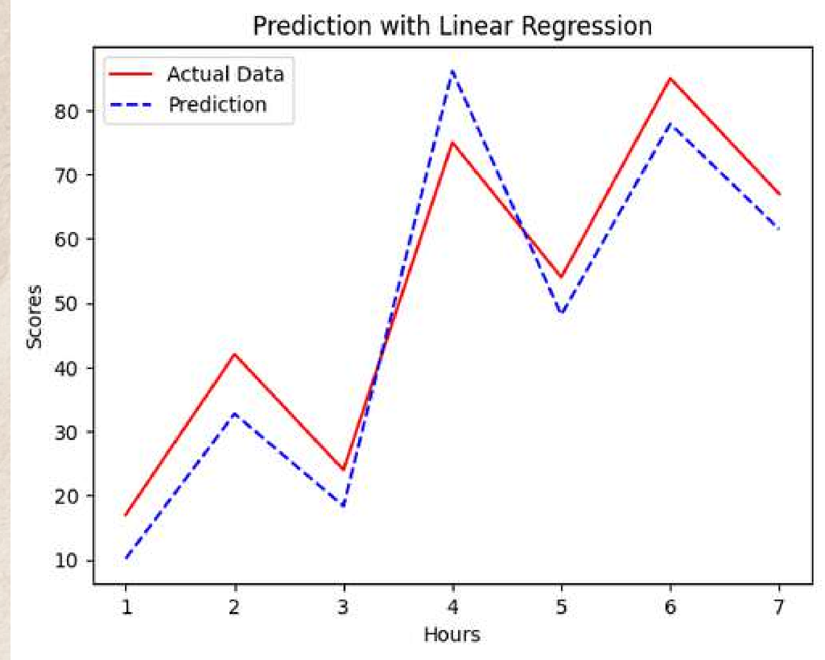
3. Random Forest Regressor

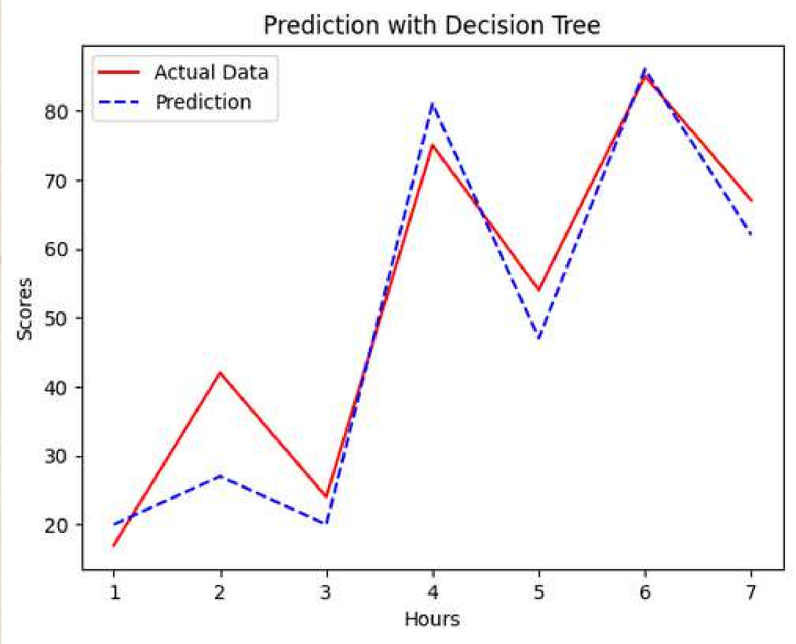
Return to TOC

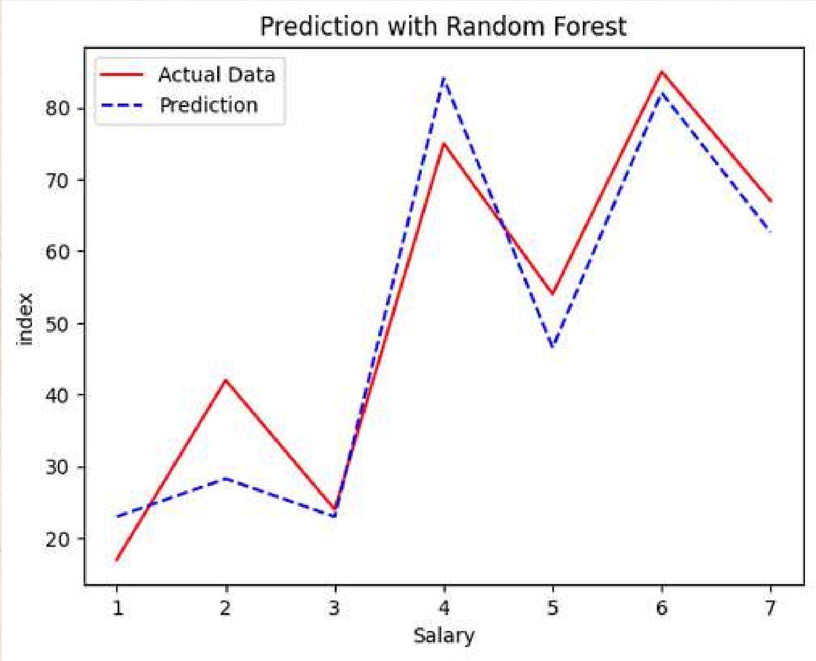
Dataset Visualization

Three regression models were implemented to predict Scores:

1. Linear Regression
   - ○ Simple and interpretable baseline model
2. Decision Tree Regressor
   - ○ Captures non-linear patterns and feature interactions
3. Random Forest Regressor
   - ○ Ensemble method to improve accuracy and reduce overfitting







Return to TOC

Model Evaluation

To compare model performance, I used Mean Absolute Error (MAE) and R^2 Score.

Model	MAE	R^2 Score
Linear Regression	5.52	0.94
Decision Tree Regressor	4.85	0.96
Random Forest Regressor	4.3	0.97

Best Performing Model: Random Forest Regressor

Return to TOC

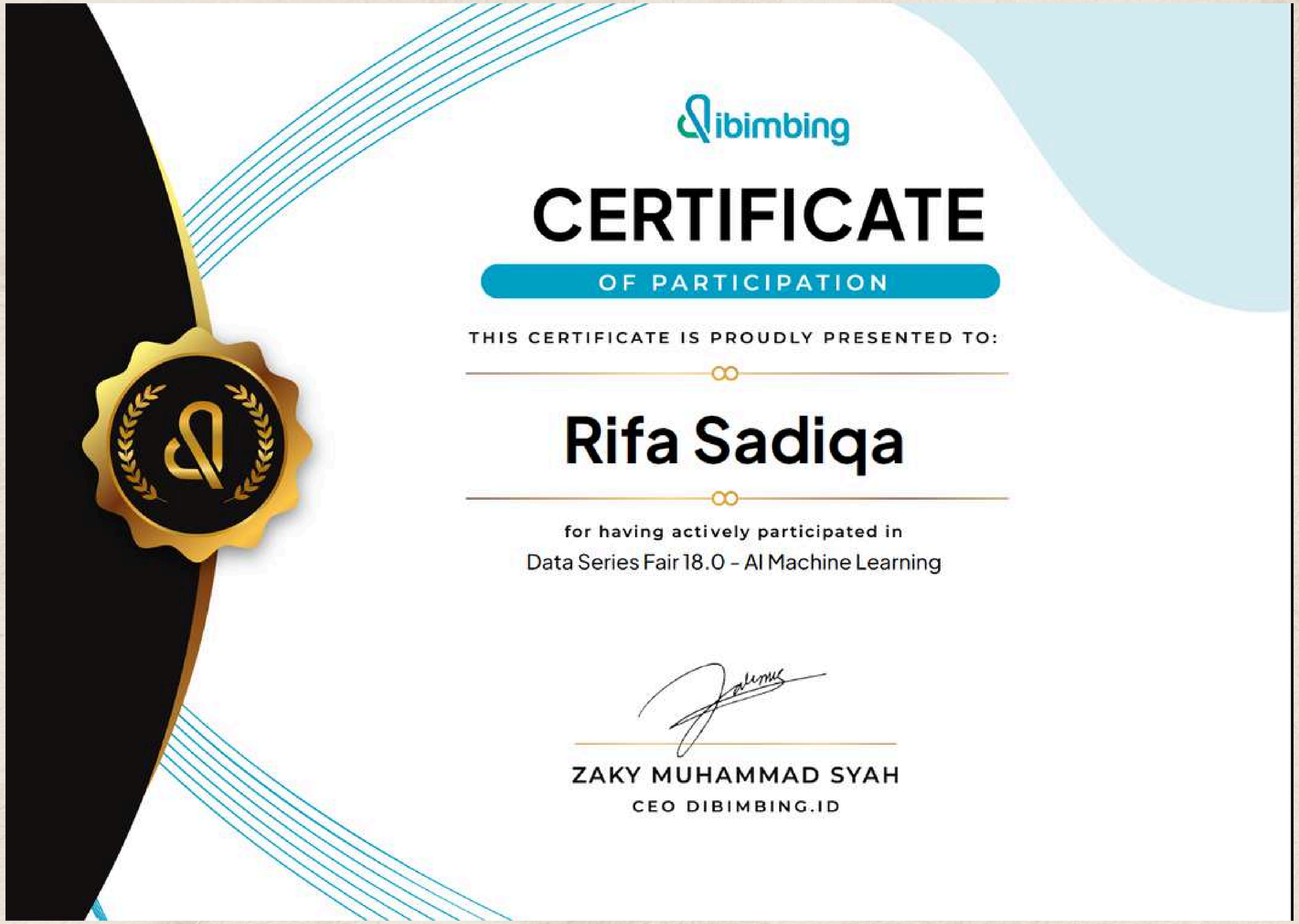
Conclusion & Takeaways

- There is a strong linear relationship between Hours Studied and Scores.
- Among all models tested, Random Forest Regressor achieved the best performance.
- This project strengthened my skills in EDA, regression modeling, and data visualization using Python.
- I'm excited to apply these skills in real-world data projects during my upcoming data science internship.

⮐ Return to TOC

dibimbing

Certificate of Completion



dibimbing

CERTIFICATE

OF PARTICIPATION

THIS CERTIFICATE IS PROUDLY PRESENTED TO:

Rifa Sadiqa

for having actively participated in

Data Series Fair 18.0 - AI Machine Learning

ZAKY MUHAMMAD SYAH

CEO DIBIMBING.ID

I am proud to have successfully completed Data Series Fair 18.0 – Data & AI Intensive Bootcamp, hosted by dibimbing.id, held on March 4–7, 2025.

This 4-day program deepened my understanding of Machine Learning, Data Science, and AI through hands-on classes and practical portfolio building.

Return to TOC

ibimbing

Thanks for Viewing - Let's Stay Connected!

linkedin.com/in/rifa-sadiqa/

github.com/rifa03

sadiqarifa12@gmail.com

rifasadiqa.my.canva.site/rifasadiqa

Return to TOC