Student Performance Prediction

Predicting Exam Scores Based on Study Hours

Prepared by: Yashi

Roll No. : [202401100400219]

Introduction

The goal of this project is to predict student exam scores based on the number of study hours and other relevant factors. We aim to understand the relationship between study efforts and academic performance through the use of data analytics and machine learning techniques. The project involves creating a predictive model using a linear regression algorithm that can estimate the exam score based on a given number of study hours.

Methodology

1. Data Collection and Preprocessing: The dataset includes two key columns: Study Hours and Exam Score. Missing values were handled by replacing them with the mean value to ensure completeness.

2. Data Visualization: A scatter plot was generated to visually represent the relationship between study hours and exam scores.

3. Model Building: A linear regression model was built with study hours as the predictor and exam score as the target. The dataset was split into training and testing sets.

4. Evaluation Metrics: The model's performance was evaluated using Mean Squared Error (MSE) and R-squared (R^2) score.

Code

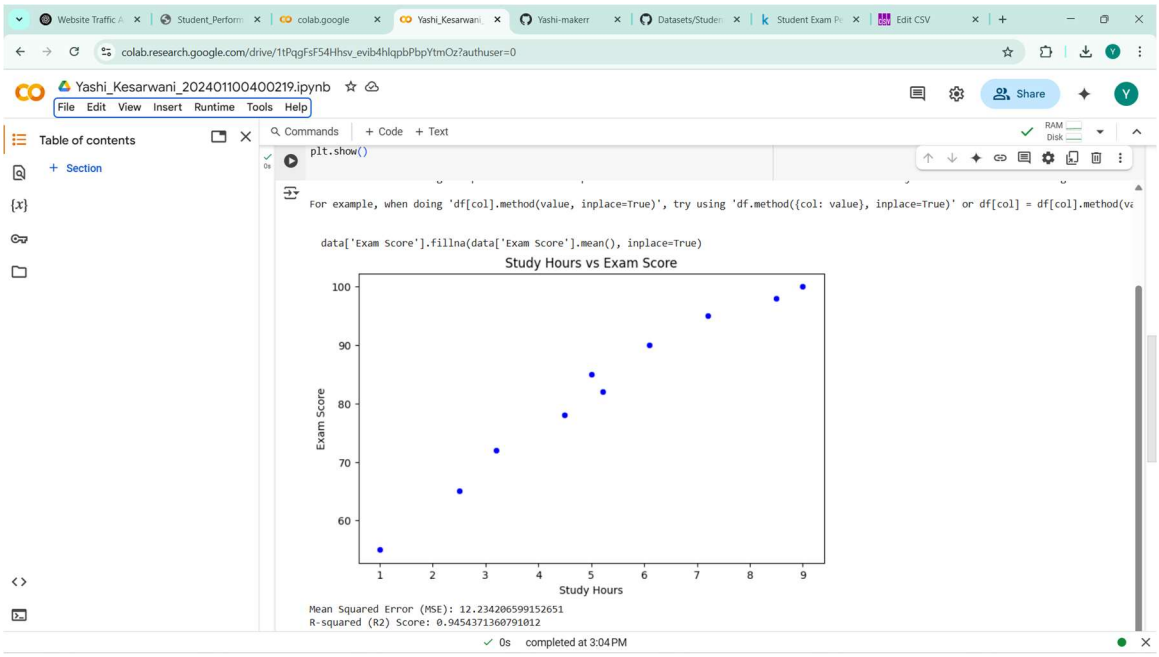
```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Load the dataset
data = pd.read_csv('YashiCSV.csv')
# Handle missing data
data['Study Hours'].fillna(data['Study Hours'].mean(), inplace=True)
data['Exam Score'].fillna(data['Exam Score'].mean(), inplace=True)

# Visualize the data
plt.figure(figsize=(8, 5))
sns.scatterplot(x=data['Study Hours'], y=data['Exam Score'])
plt.title('Study Hours vs Exam Score')
plt.show()

# Build and train the linear regression model
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

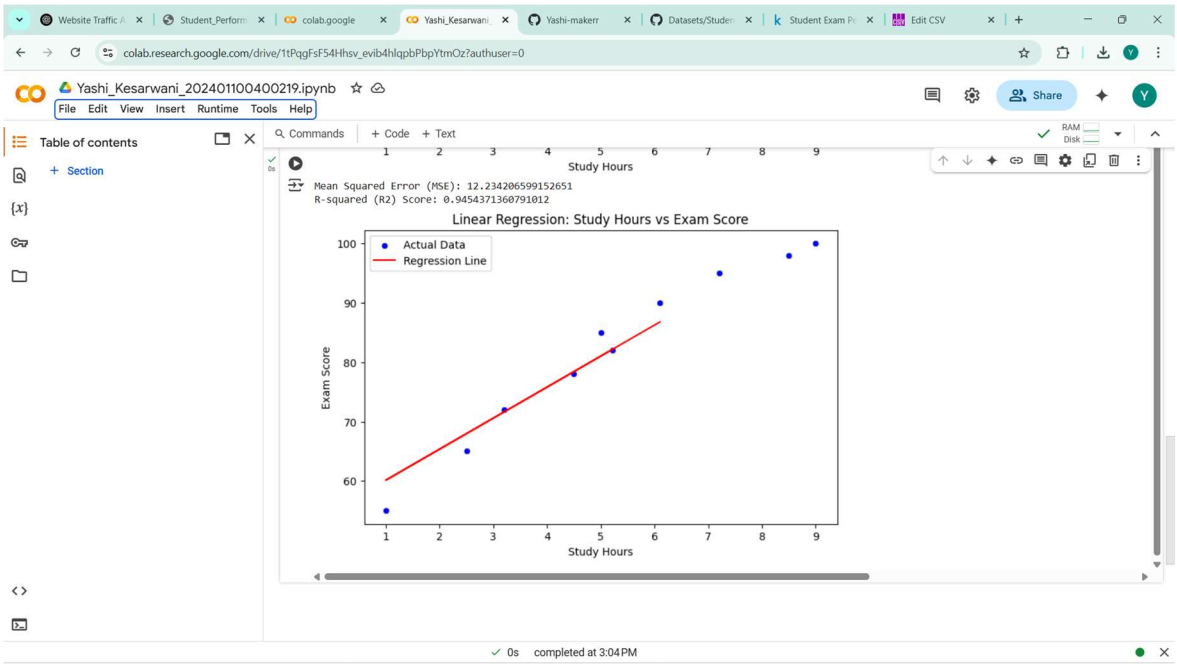
Output/Result

The output of the code includes:

1. A scatter plot visualizing the relationship between study hours and exam scores.
2. A regression line showing the model's prediction.
3. The MSE and R^2 score as evaluation metrics.

Screenshot of the results can be attached separately.





References/Credits

1. Dataset provided for academic purposes.
2. Libraries used: Pandas, Seaborn, Matplotlib, Scikit-learn.
3. Linear Regression model implemented using Scikit-learn.