Student Performance Prediction

Name: [Abhishek Kumar Maurya]

Roll No: [08]

Introduction

This report aims to predict student performance based on study hours and previous scores. The dataset contains information about 20 students, including their study hours, previous scores, and final exam scores. The objective is to analyze the relationship between these factors and predict final scores based on given inputs.

Methodology

1. Data Collection: The dataset contains 20 records with attributes: StudentID, StudyHours, PreviousScores, and FinalExamScore.

2. Preprocessing:
  - Checked for missing values (None found).
  - Ensured all data types were appropriate for analysis.

3. Analysis Approach:
  - Examined correlations between study hours, previous scores, and final exam scores.
  - Used a simple linear regression model to predict performance based on StudyHours and PreviousScores.

Code

```
# Import necessary libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns


# Load the dataset from uploaded CSV file

file_path = '/content/student_data.csv'

df = pd.read_csv(file_path)



# Basic statistics of the dataset
```

```
print("\nBasic Statistics:")

print(df.describe())



# Splitting data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Display basic information about the dataset

print("Dataset Overview:")

print(df.info())

print("\nFirst 5 Rows:")

print(df.head())



# Data Visualization

plt.figure(figsize=(10, 6))

sns.histplot(df[df.columns[1]], bins=20, kde=True, color='blue')  # Assuming first numeric column

plt.title(f'Distribution of {df.columns[1]}')

plt.xlabel(df.columns[1])

plt.ylabel('Frequency')

plt.show()
```

Output/Result

• A linear regression model was trained to predict final exam scores.
• The Mean Absolute Error (MAE) of the model was calculated to assess performance.
• The analysis revealed that both study hours and previous scores significantly impact final exam scores.

References/Credits

• Dataset provided by [Data Provided student_data.csv]

• Python libraries used: Pandas, mat-plotlib.

GitHub Repository

[Insert your repository link here]