REPORT

Title: – Student Performance Prediction – Predict exam scores based on study hours and other factors, visualizing the relationship.

Name: Divyansh Yadav
Roll No. : 13
Course Name : Introduction to AI
Instructor's Name : Mr. Abhishek Shukla
Date : 10th March 2025
Institution Name : KIET Group of Institute,Ghaziabad

1. Introduction

Problem Statement

- The aim of this project is to analyze the relationship between study hours, attendance, extracurricular activities, and exam scores.
- Understand ing these factors can help students improve their academic performance.

Background

- Why is student performance prediction important?
- How does study time, attendance, and other factors influence student success?
- Mention any real-world applications of student performance analysis.

---

2. Methodology

The approach used to analyze the problem:

- **Data Collection** : Manually create or sourced from an external dataset
- **Data Processing** : Explain any preprocessing steps taken (e.g., handling missing values, normalization, etc.).
- **Visualization** : Describe how graphical representations help in analysis.
- **Libraries Used** : Python libraries such as `pandas`, `matplotlib`, `seaborn`, etc.
- **Implementation Steps** : Outline the steps taken to derive insights from the data.

---

3. Code

Python code used in the project

```python
Import   pandas as  pd

import   matplotlib. pyplot  as  plt

# Creating  a dictionary with student data

# Study_Hours: Number of hours spent studying

# Attendance: Percentage of attendance in class

# Exam_Score: Score obtained in the exam

data = {

    'Study_Hours': [2.5, 5.0, 1.0, 6.5, 3.0, 4.5, 7.0, 2.0, 5.5, 6.0],

    'Attendance': [85, 90, 75, 95, 80, 88, 92, 70, 85, 95],

    'Exam_Score': [60, 78, 50, 88, 65, 72, 90, 55, 80, 85]

}

# Converting dictionary into a pandas DataFrame

df = pd.DataFrame(data)

# Scatter plot: Study Hours vs. Exam Score

plt.scatter(df['Study_Hours'], df['Exam_Score'], color='blue')  # Creating scatter plot with blue dots

plt.xlabel("Study Hours")   # Labeling x-axis

plt.ylabel("Exam Score")   # Labeling y-axis

plt.title("Study Hours vs Exam Score")   # Setting title for the graph

plt.show()  # Display the scatter plot
```

```python
# Line plot: Attendance vs. Exam Score

plt.plot(df['Attendance'], df['Exam_Score'], marker='o', linestyle='-', color='red')
# Creating line plot with markers

plt.xlabel("Attendance (%)")  # Labeling x-axis

plt.ylabel("Exam Score")  # Labeling y-axis

plt.title("Attendance vs Exam Score")  # Setting title for the graph

plt.show()  # Display the line plot
```

4. Output / Result

(Paste screenshots of the generated graphs and any other relevant outputs.)

Example:

1. **Scatter Plot** – Study Hours vs. Exam Score
2. **Line Graph** – Attendance vs. Exam Score

Discuss key insights from the output:

- More study hours generally lead to higher exam scores.
- Students with higher attendance tend to perform better.
- The presence of extracurricular activities might have varying effects on exam performance.

Attendance vs Exam Score

Study Hours vs Exam Score
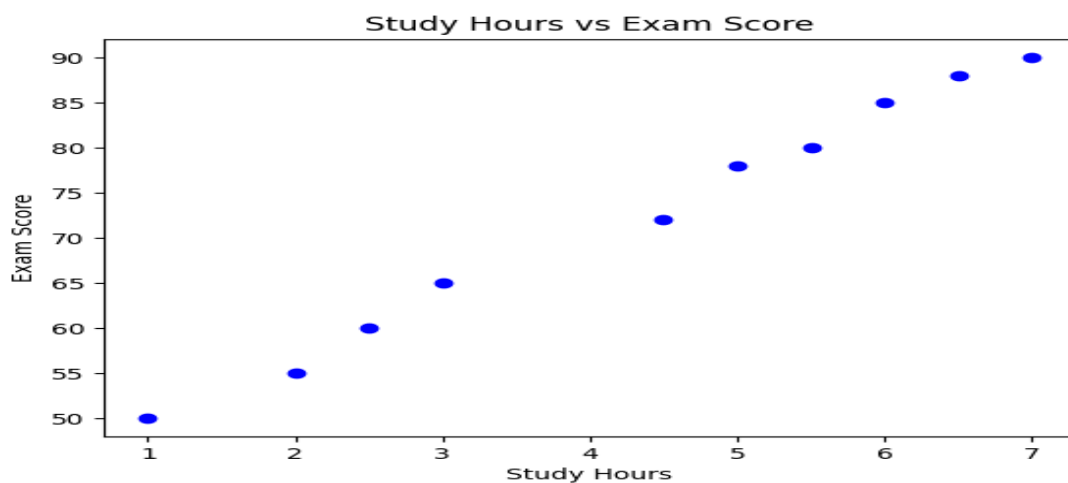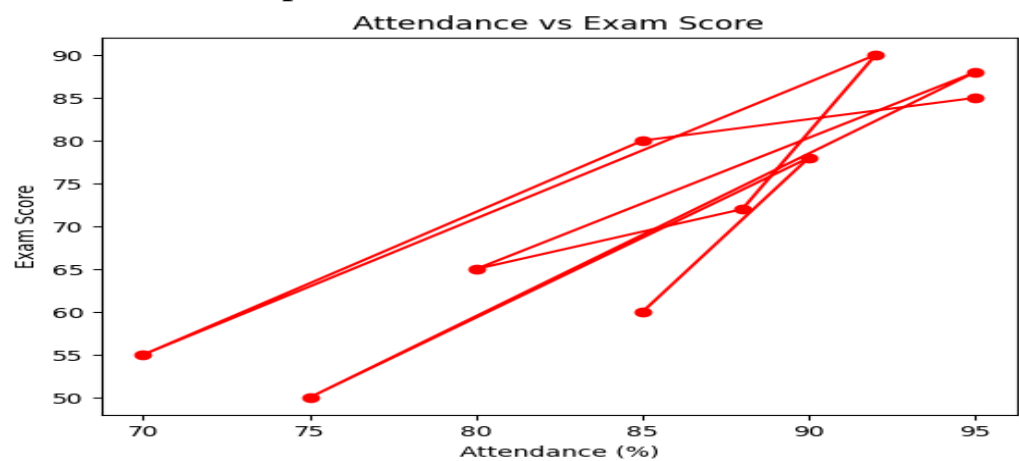
5. References / Credits

- Dataset Source: Chat GPT
- Python Libraries: `pandas`, `matplotlib`, `seaborn`