AI FINAL PROJECT REPORT

SUBMITTED TO: SIR ADNAN KARAMAT

SUBMITTED FROM: MOMIN SAQIB NIAZI (BAI243073)

AI Project Report: Salary Prediction using Linear Regression

1. Project Overview

The objective of this project is to predict an employee's salary based on their years of professional experience. This is a **Simple Linear Regression** task where we use one independent variable to predict one continuous dependent variable.

2. Dataset Description

- **Source:** Kaggle (Salary Dataset).

- **Features (X):** YearsExperience (Numerical).

- **Target (y):** Salary (Numerical).

- **Size:** 30 entries (split into 80% Training and 20% Testing).

3. Methodology

- **Algorithm:** Linear Regression.

- **Tools:** Python, Pandas, Scikit-Learn, Matplotlib.

- **Implementation:** The data was split using train_test_split. The model was trained using the LinearRegression() class from Scikit-Learn.

4. Results & Performance

Based on the implementation in Google Colab, the model achieved the following performance:

- **R^2 Score (Accuracy): 0.9024**

- **Interpretation:** An R^2 score of ~0.90 means that approximately **90% of the variance** in salary can be explained by the years of experience. This indicates a very high level of accuracy for the model.

5. Conclusion

The visual representation shows a strong positive correlation. The blue regression line closely follows the actual red data points, proving that as experience increases, salary increases in a predictable linear fashion.

CSV FILE:

	A	B	C	D	E	F	G	H
1		YearsExpe	Salary					
2	0	1.2	39344					
3	1	1.4	46206					
4	2	1.6	37732					
5	3	2.1	43526					
6	4	2.3	39892					
7	5	3	56643					
8	6	3.1	60151					
9	7	3.3	54446					
10	8	3.3	64446					
11	9	3.8	57190					
12	10	4	63219					
13	11	4.1	55795					
14	12	4.1	56958					
15	13	4.2	57082					
16	14	4.6	61112					
17	15	5	67939					
18	16	5.2	66030					
19	17	5.4	83089					
20	18	6	81364					
21	19	6.1	93941					
22	20	6.9	91739					
23	21	7.2	98274					
24	22	8	101303					
25	23	8.3	113813					
26	24	8.8	109432					
27	25	9.1	105583					
28	26	9.6	116970					

Salary_dataset   +

CODE:

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score


# Load and clean data
df = pd.read_csv('Salary_dataset.csv')
X = df[['YearsExperience']]
y = df['Salary']


# Split, Train, and Predict
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)


# Show results
print(f"R2 Score: {r2_score(y_test, y_pred)}")
plt.scatter(X_test, y_test, color='red', label='Actual')
plt.plot(X_test, y_pred, color='blue', label='Predicted')
plt.legend()
plt.show()
```