Madhav Institute of Technology & Science, Gwalior (M.P.)

(A Govt. Aided UGC Autonomous Institute & NAAC Accredited, Estd. in 1957, Affiliated to RGPV Bhopal)

SKILL-BASED PROJECT

SEMESTER: 6TH

SUBJECT: AIML

SUBMITTED BY:
PURAN PATEL
0901AU201021

SUBMITTED TO:
PROF. VAIBHAV SHIVHARE
PROF. AJAY SINGH RAJPUT

0901AU201021 Puran Patel

AIML Skill Based Project

Salary Prediction ML Model

Importing Libraries

In [2]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Loading the Dataset

In [3]:
```
df = pd.read_csv('Salary Data.csv')

df.head(11)
```

Out[3]:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32	Male	Bachelor's	Software Engineer	5.0	90000
1	28	Female	Master's	Data Analyst	3.0	65000
2	45	Male	PhD	Senior Manager	15.0	150000
3	36	Female	Bachelor's	Sales Associate	7.0	60000
4	52	Male	Master's	Director	20.0	200000
5	29	Male	Bachelor's	Marketing Analyst	2.0	55000
6	42	Female	Master's	Product Manager	12.0	120000
7	31	Male	Bachelor's	Sales Manager	4.0	80000
8	26	Female	Bachelor's	Marketing Coordinator	1.0	45000
9	38	Male	PhD	Senior Scientist	10.0	110000
10	29	Male	Master's	Software Developer	3.0	75000

In [4]:
```
df.tail(11)
```

Out[4]:

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
362	35	Female	Bachelor's	Senior Financial Manager	8.0	90000
363	43	Male	Master's	Director of Marketing	18.0	170000
364	31	Female	Bachelor's	Junior Financial Analyst	3.0	50000
365	41	Male	Bachelor's	Senior Product Manager	14.0	150000
366	44	Female	PhD	Senior Data Engineer	16.0	160000
367	33	Male	Bachelor's	Junior Business Analyst	4.0	60000
368	35	Female	Bachelor's	Senior Marketing Analyst	8.0	85000
369	43	Male	Master's	Director of Operations	19.0	170000
370	29	Female	Bachelor's	Junior Project Manager	2.0	40000
371	34	Male	Bachelor's	Senior Operations Coordinator	7.0	90000
372	44	Female	PhD	Senior Business Analyst	15.0	150000

Splitting the dataset

In [5]:
```
X = df[['Age']]
y = df[['Salary']]
X_train, X_test, y_train, y_test = train_test_split(X.values, y, test_size=0.2, random_state=42)
```

Creating a model

In [6]:
```
model = LinearRegression()
model.fit(X_train, y_train)
```

Out[6]:
```
▾ LinearRegression
LinearRegression()
```

Evaluating the model

In [7]:
```
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print("Mean Squared Error:", mse)
print("R-Squared Value:", r2)
```

```
Mean Squared Error: 362326711.91360664
R-Squared Value: 0.8488782955782094
```

Predicting salary for given age

In [8]:
```
# 1st Prediction
age = [[25]]
income = model.predict(age)
print("Predicted Income for age", age, ":", income)
```

```
Predicted Income for age [[25]] : [[21518.89225188]]
```

In [9]:
```
# 2nd Prediction
age = [[30]]
income = model.predict(age)
print("Predicted Income for age", age, ":", income)
```

```
Predicted Income for age [[30]] : [[52977.40932461]]
```

Visualizing the training set results

In [11]:
```
import matplotlib.pyplot as plt

plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, model.predict(X_train), color = 'blue')
plt.title('Salary vs Age (Training set)')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()
```

Visualizing the test set results

In [12]:
```
plt.scatter(X_test, y_test, color = 'red')
plt.plot(X_train, model.predict(X_train), color = 'blue')
plt.title('Salary vs Age (Test set)')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()
```

Final Linear Regression Equation for model with coefficient values

In [13]:
```
print(model.coef_)
print(model.intercept_)
```

```
[[6291.70341455]]
[-135773.69311178]
```

Hence, the model equation is given as:

Salary = (6291.7034 * Age) - 135773.6931

Reference:

The Data Set is taken from Kaggle website. The link for the dataset:
https://www.kaggle.com/datasets/rkiattisak/salaly-prediction-for-beginer

In [ ]: