

Assignment 1 – Linear Regression

Problem Statement:

You work in XYZ Company as a Python Data Scientist. The company officials have collected some data on salaries based on year of experience and wish for you to create a model from it. Dataset: data.csv

Tasks To Be Performed:

1. Load the dataset using pandas
2. Extract data from years experience column is a variable named X
3. Extract data from the salary column is a variable named Y
4. Divide the dataset into two parts for training and testing in 66% and 33% proportion
5. Create and train Linear Regression Model on training set
6. Make predictions based on the testing set using the trained model
7. Check the performance by calculating the r2 score of the model

In [1]:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import *
```

In [3]:

```
df = pd.read_csv(r"csv files/data.csv")
df.head()
```

Out[3]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

In [8]:

```
X = df[['YearsExperience']]
y = df[['Salary']]
```

In [9]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.66, random_state=0)
```

In [10]:

```
lr = LinearRegression()
lr.fit(X_train, y_train)
```

Out[10]:

▼ LinearRegression

LinearRegression()

In [11]:

```
y_pred = lr.predict(X_test)
```

In [12]:

```
r2_score(y_test, y_pred)
```

Out[12]:

0.9766870911747516

Model is 97% efficient

In []:

