

Hiring Salary Prediction

A company hires candidates based on:

- Years of experience
- Written test score (out of 10)
- Interview score (out of 10)

The HR department wants to automate the salary prediction process using machine learning. This project aims to build a Linear Regression model that will predict a candidate's salary based on these three factors.

Objective

Build and train a Linear Regression model using historical hiring data and use it to predict salaries for the following candidates:

1. Candidate A: 2 years experience, 9 test score, 6 interview score
2. Candidate B: 12 years experience, 10 test score, 10 interview score

```
In [8]: import pandas as pd
import numpy as np
from sklearn import linear_model
from word2number import w2n
```

```
In [15]: # Importing Data
Hiring_data = pd.read_csv("/Users/priyanshi/Downloads/hiring.csv")
Hiring_data
```

Out [15]:

	experience	test_score(out of 10)	interview_score(out of 10)	salary(\$)
0	NaN	8.0	9	50000
1	NaN	8.0	6	45000
2	five	6.0	7	60000
3	two	10.0	10	65000
4	seven	9.0	6	70000
5	three	7.0	10	62000
6	ten	NaN	7	72000
7	eleven	7.0	8	80000

In [16]: *#Filling Missing value*
Hiring_data.experience = Hiring_data.experience.fillna("zero")
Hiring_data

Out [16]:

	experience	test_score(out of 10)	interview_score(out of 10)	salary(\$)
0	zero	8.0	9	50000
1	zero	8.0	6	45000
2	five	6.0	7	60000
3	two	10.0	10	65000
4	seven	9.0	6	70000
5	three	7.0	10	62000
6	ten	NaN	7	72000
7	eleven	7.0	8	80000

```
In [17]: # converting word to number
Hiring_data.experience = Hiring_data.experience.apply(w2n.word_to_num)
Hiring_data
```

```
Out[17]:
```

	experience	test_score(out of 10)	interview_score(out of 10)	salary(\$)
0	0	8.0	9	50000
1	0	8.0	6	45000
2	5	6.0	7	60000
3	2	10.0	10	65000
4	7	9.0	6	70000
5	3	7.0	10	62000
6	10	NaN	7	72000
7	11	7.0	8	80000

```
In [22]: import math
mean_test_score = math.floor(Hiring_data['test_score(out of 10)'].mean())
mean_test_score
```

```
Out[22]: 7
```

```
In [23]: Hiring_data['test_score(out of 10)'] = Hiring_data['test_score(out of 10)'].fillna(mean_test_score)
Hiring_data
```

Out [23]:

	experience	test_score(out of 10)	interview_score(out of 10)	salary(\$)
0	0	8.0	9	50000
1	0	8.0	6	45000
2	5	6.0	7	60000
3	2	10.0	10	65000
4	7	9.0	6	70000
5	3	7.0	10	62000
6	10	7.0	7	72000
7	11	7.0	8	80000

In [31]:

```
# Define X and y

X = Hiring_data[['experience','test_score(out of 10)','interview_score(out of 10)']] # Independent variable
y = Hiring_data['salary($)'] # Dependent variable

# Train the model
reg = linear_model.LinearRegression()
reg.fit(X, y)

# Convert DataFrame to NumPy array before training
reg.fit(X.values, y)
```

Out [31]:

▼ LinearRegression ⓘ ?

LinearRegression()

Candidate A: 2 years experience, 9 test score, 6 interview score Predicted Salary : 53713

In [32]:

```
reg.predict([[2,9,6]])
```

Out [32]: array([53713.86677124])

Candidate B: 12 years experience, 10 test score, 10 interview score Predicted Salary:93747

In [34]: reg.predict([[12,10,10]])

Out [34]: array([93747.79628651])

In []: