1. **import pandas as pd** (Untuk mengimport library pandas kemudian diinisialisasi dengan varibel pd)
2. **from sqlalchemy import create\_engine** (Dari sqlalchemy diimport metode create\_engine)
3. **import pymysql** (Untuk mengimport library pymysql)
4. **sqlEngine = create\_engine('mysql+pymysql://root:@127.0.0.1', pool\_recycle=3600)** (Membuat engine database MySQL kemudian disimpan dalam variabel sqlEngine)
5. **dbConnection = sqlEngine.connect()** (Membuat koneksi dengan metode sqlEngine agar terhubung dengan database, kemudian simpan dalam variabel dbConnection)
6. **df = pd.read\_sql('SELECT \* FROM djpegawai.prediksi\_pegawai ', dbConnection)** (Membaca data pada database djpegawai tabel prediksi\_pegawai dan dikoneksikan dengan dbConnections, kemudian disimpan pada variabel df)
7. **pd.set\_option('display.expand\_frame\_repr', False)** ()
8. **print(df)** (Menampilkan data pada variabel df)
9. **dbConnection.close()** ()
10. **df.head()** (Untuk menampilkan 5 data teratas yang terdapat dalam variabel df)
11. **df** (Untuk menampilkan seluruh data yang terdapat dalam variabel df)
12. **df.tail()** (Untuk menampilkan 5 data terakhir yang terdapat dalam variabel df)
13. **df.info()** (Untuk menampilkan struktur data)
14. **x = df.iloc[:, :-1]** (Membuat variabel X yang di dalamnya column yang diambil itu column Salary)
15. **y = df.iloc[:, 1]** (Membuat variabel Y yang di dalamnya column yang diambil itu column YearsExperience)
16. **from sklearn.model\_selection import train\_test\_split** (Dari scikit learn yang dimana menggunakan model selection untuk mengimport metode train\_test\_split)
17. **x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=1/3, random\_state=0)** (Membuat variabel x\_train, x\_test, y\_train, dan y\_test menggunakan metode train\_test\_split yang dimana variabel yang digunakan adalah x, y, dan untuk data testnya sebanyak 1/3 total data yang ada dengan pemilihan data secara acak)
18. **from sklearn.linear\_model import LinearRegression** (Dari scikit learn linear model diimport metode LinearRegression)
19. **regressor = LinearRegression()** (Membuat variabel regressor yang isinya metode LinearRegression)
20. **equation = regressor.fit(x\_train, y\_train)** (Membuat equation yang isinya terdapat method regressor.fit dengan parameternya x\_train dan y\_train)
21. **print(regressor.coef\_)** (Mencari koefisien dari x menggunakan regressor, kemudian ditampilkan)
22. **print(regressor.intercept\_)** (Mencari konstanta/intercept menggunakan regressor, kemudian ditampilkan)
23. **y\_pred = equation.predict(x\_test)** (Menentukan variabel y\_pred yang didalamnya mencari prediksi mengggunakan variabel equation dengan parameternya x\_test)
24. **import matplotlib.pyplot as plt** (Mengimport library matplotlib.pyplot)
25. **plt.scatter(x\_train, y\_train, color="green")** (Menampilkan titik/scatter x\_train dan y\_train berwarna hijau)
26. **plt.title("Salary Prediction")** (Membuat judul grafik dengan nama Salary Prediction)
27. **plt.xlabel("Years Experience")** (Membuat label sumbu x dengan nama Years Experience)
28. **plt.ylabel("Salary")** (Membuat label sumbu y dengan nama Salary)
29. **plt.grid()** (Membuat garis pada grafik)
30. **plt.show()** (Menampilkan visualisasi grafik)
31. **import matplotlib.pyplot as plt** (Mengimport library matplotlib.pyplot)
32. **plt.scatter(x\_train, y\_train, color="green")** (Menampilkan titik/scatter x\_train dan y\_train berwarna hijau)
33. **plt.plot(x\_train, regressor.predict(x\_train), color="red")** (Menampilkan plot/garis pada grafik berwarna merah)
34. **plt.title("Salary Prediction")** (Membuat judul grafik dengan nama Salary Prediction)
35. **plt.xlabel("Years Experience")** (Membuat label sumbu x dengan nama Years Experience)
36. **plt.ylabel("Salary")** (Membuat label sumbu y dengan nama Salary)
37. **plt.legend(('$\hat{y}(x)=9345.94x+26816.19$',),loc='upper right')** (Menampilkan fungsi linear y = 9345.94x + 26816.19 pada grafik)
38. **plt.grid()** (Membuat garis pada grafik)
39. **plt.show()** (Menampilkan visualisasi grafik)
40. **from sklearn.metrics import r2\_score** (Dari scikit learn metrics diimport metode r2\_score)
41. **r2 =r2\_score (x\_train,y\_train)** (Menghitung r2\_score dengan parameternya x\_train dan y\_train, kemudian hasilnya disimpan pada variable r2)
42. **print ('Nilai R2 adalah',r2)** (Menampilkan nilai pada variabel r2)
43. **import numpy as np** (Untuk mengimport libraty numpy kemudian diinisialisasi dengan variabel np)
44. **import pandas as pd** (Untuk mengimpoty library pandas kemudian diinisialisasi dengan variabel pd)
45. **import scipy** (Untuk mengimport library scipy)
46. **from scipy import stats** (Dari Scipy diimport metode stats)
47. **def ANOVATAB(y,yhat,n,m):**

**dfn=n**

**dfd=m-n-1**

**ybar=np.average(y)**

**SSR=sum((yhat-ybar)\*\*2)**

**SSE=sum((y-yhat)\*\*2)**

**print (SSE)**

**print (dfd)**

**SST=sum((y-ybar)\*\*2)**

**MSR=SSR/dfn**

**MSE=SSE/dfd**

**Fs=MSR/MSE**

**ks =1-scipy.stats.f.cdf(Fs,dfn,dfd)**

**data\_table={**

**'SS' : [SSR,SSE,SST],**

**'df' : [dfn,dfd,m-1],**

**'MS' : [MSR,MSE,'-'],**

**'Fs' : [Fs,'-','-'],**

**'pval' : [ks,'-','-']**

**}**

**return pd.DataFrame(data\_table)**

1. **n=1**

**m=len(y\_pred)**

**print('Tabel Anovanya')**

**result=ANOVATAB(y\_test,y\_pred,n,m)**

**result**

1. **plt.scatter(x\_test, y\_test, color="green")** (Menampilkan titik/scatter x\_test dan y\_test berwarna hijau)
2. **plt.title("Salary Prediction")** (Membuat judul grafik dengan nama Salary Prediction)
3. **plt.xlabel("Years Experience")** (Membuat label sumbu x dengan nama Years Experience)
4. **plt.ylabel("Salary")** (Membuat label sumbu y dengan nama Salary)
5. **plt.grid()** (Membuat garis pada grafik)
6. **plt.show()** (Menampilkan visualisasi grafik)
7. **plt.scatter(x\_test, y\_test, color="green")** (Menampilkan titik/scatter x\_test dan y\_test berwarna hijau)
8. **plt.plot(x\_train, regressor.predict(x\_train), color="red")** (Menampilkan plot/garis pada grafik berwarna merah)
9. **plt.title("Salary Prediction")** (Membuat judul grafik dengan nama Salary Prediction)
10. **plt.xlabel("Years Experience")** (Membuat label sumbu x dengan nama Years Experience)
11. **plt.ylabel("Salary")** (Membuat label sumbu y dengan nama Salary)
12. **plt.grid()** (Membuat garis pada grafik)
13. **plt.show()** (Menampilkan visualisasi grafik)
14. **salary\_pred = regressor.predict([[15.5]])** (Menentukan prediksi gaji dengan parameter YearsExperience nya 15.5 tahun menggunakan metode pada variabel regressor, kemudian simpan di variabel salary\_pred)
15. **print("The salary for that amount of years experience is: ", salary\_pred)** (Menampilkan nilai dari variabel salary\_pred)
16. **import pickle** (Untuk import library pickle)
17. **pd.to\_pickle(equation, "model3.pickle")** (Untuk mengconvert model machine learning ke dalam file model3.pickle)