

Nama : Dimas Gary Irawan

Metode Numerik – D

NIM : 21120122140164

Link Github : https://github.com/garyirawan/Aplikasi-Regresi_Dimas-Gary-Irawan_21120122140164

APLIKASI REGRESI UNTUK PEMECAHAN PROBLEM

Tugas mahasiswa:

1. Mahasiswa membuat **kode sumber** dengan bahasa pemrograman yang dikuasai untuk mengimplementasikan solusi di atas, dengan ketentuan:
 - o NIM terakhir % 4 = 0 mengerjakan Problem 1 dengan Metode 1 dan Metode 2
 - o NIM terakhir % 4 = 1 mengerjakan Problem 1 dengan Metode 1 dan Metode 3
 - o NIM terakhir % 4 = 2 mengerjakan Problem 2 dengan Metode 1 dan Metode 2
 - o NIM terakhir % 4 = 3 mengerjakan Problem 2 dengan Metode 1 dan Metode 3
 - o Mahasiswa juga bisa menambah solusi dengan salah satu metode opsional

Berdasarkan dari syarat yang diminta, saya harus mengerjakan Problem 1 dengan Metode 1 dan Metode 2. $4 \% 4 = 0$.

Diinginkan untuk mencari hubungan faktor yang mempengaruhi nilai ujian siswa (NT):

1. Durasi waktu belajar (TB) terhadap nilai ujian (Problem 1)

Implementasikan regresi untuk mencari hubungan tersebut menggunakan metode:

1. Model linear (Metode 1)
2. Model pangkat sederhana (Metode 2)

Hitung galat RMS dari tiap metode yang digunakan.

REGRESI LINEAR

Regresi Linear adalah metode statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (prediktor) dan satu variabel dependen (respon). Tujuan dari regresi linear adalah menemukan garis yang paling cocok (disebut garis regresi) yang memprediksi nilai variabel dependen berdasarkan nilai variabel independen. Untuk regresi linear sederhana (dengan satu variabel independen), modelnya adalah:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Di mana:

y adalah variabel dependen (target).

x adalah variabel independen (prediktor).

β_0 adalah intercept (titik di mana garis memotong sumbu y).

β_1 adalah koefisien kemiringan (slope) dari garis regresi.

ϵ adalah error atau residu.

Implementasi Kode:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from scipy.optimize import curve_fit

url = 'C:/Users/gary/python/Numerik-
Regresi/student_performance.csv'
data = pd.read_csv(url)
print(data.head())

X = data['Hours Studied'].values.reshape(-1, 1)
y = data['Performance Index'].values

linear_model = LinearRegression()
linear_model.fit(X, y)
y_pred_linear = linear_model.predict(X)

rms_error_linear = np.sqrt(mean_squared_error(y, y_pred_linear))

plt.scatter(X, y, color='blue', label='Data Asli')
plt.plot(X, y_pred_linear, color='red', label='Regresi Linear')
plt.xlabel('Hours Studied')
plt.ylabel('Performance Index')
plt.title('Regresi Linear')
plt.legend()
plt.show()

print('RMS Error untuk Model Linear:', rms_error_linear)
```

Ringkasan

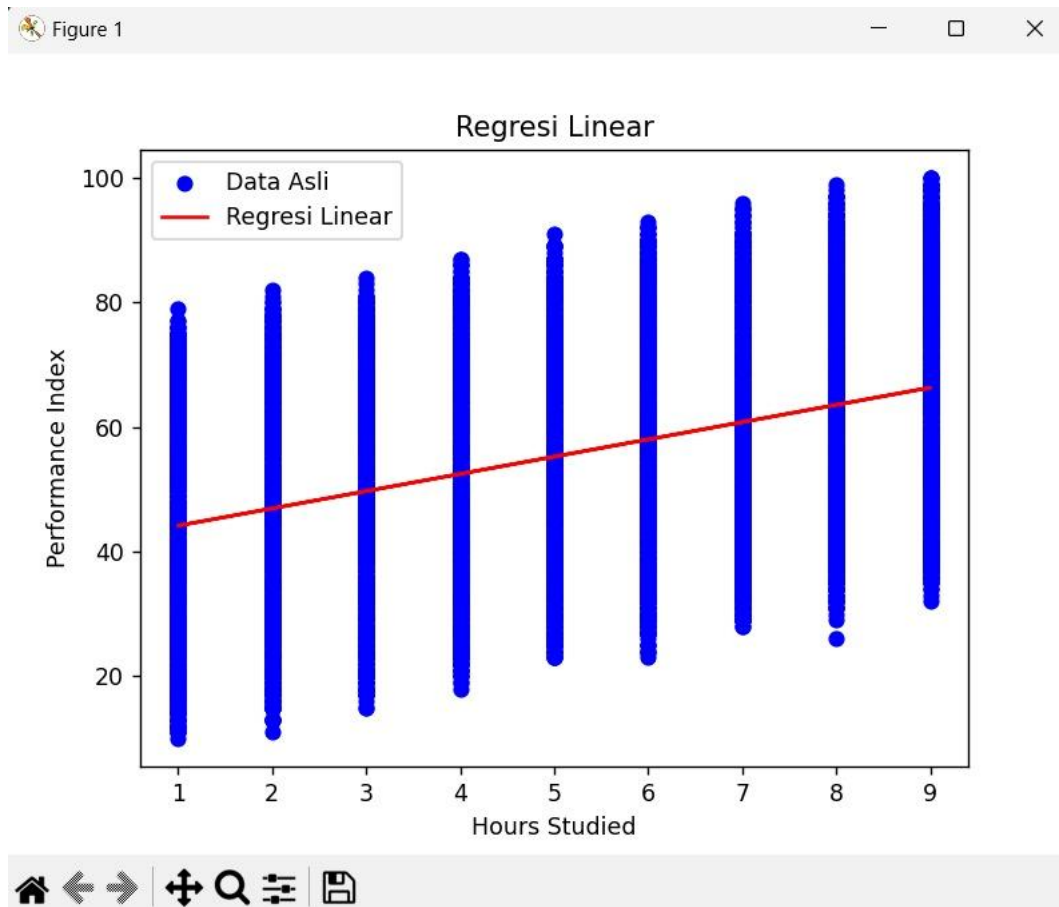
Kode di atas melakukan analisis regresi linear sederhana pada data yang terdiri dari dua variabel: 'Hours Studied' (jam belajar) dan 'Performance Index' (nilai ujian). Tujuan utamanya adalah memodelkan hubungan antara jam belajar dan nilai ujian menggunakan regresi linear dan mengevaluasi kesalahan model menggunakan Root Mean Squared Error (RMSE).

Konsep

1. Regresi Linear: Teknik statistik untuk memodelkan hubungan antara variabel dependen (y) dan satu atau lebih variabel independen (X) dengan menyesuaikan garis lurus (model linear) ke data yang diamati.
2. Root Mean Squared Error (RMSE): Ukuran kesalahan antara nilai yang diprediksi oleh model dan nilai sebenarnya. RMSE memberikan gambaran tentang seberapa baik model memprediksi data yang sebenarnya.
3. Visualisasi: Menggunakan scatter plot untuk menampilkan data asli dan garis regresi untuk menunjukkan hasil model linear.

Hasil Pnegujian

```
C:\Users\gary\python>C:/Users/gary/AppData/Local/Programs/Python/Python311/
python.exe c:/Users/gary/python/Numerik-Regresi/Linear.py
Hours Studied ... Performance Index
0          7 ...          91.0
1          4 ...          65.0
2          8 ...          45.0
3          5 ...          36.0
4          7 ...          66.0
```



Analisis Hasil

1. Hubungan Linear: Scatter plot dan garis regresi menunjukkan bahwa ada hubungan linear antara 'Hours Studied' dan 'Performance Index'. Hal ini berarti semakin banyak jam belajar, semakin tinggi nilai ujian, yang konsisten dengan ekspektasi umum.
2. Kesalahan Model: RMSE memberikan gambaran tentang seberapa baik model melakukan prediksi. Nilai RMSE yang lebih rendah menunjukkan bahwa model memiliki prediksi yang lebih akurat. Dalam kode ini, nilai RMSE yang dihasilkan memberikan indikasi seberapa besar deviasi antara nilai yang diprediksi dan nilai sebenarnya.
3. Kekuatan Model: Garis regresi yang digambarkan pada plot menunjukkan bahwa model linear cukup baik dalam menangkap tren umum dalam data,

meskipun mungkin ada beberapa outlier atau deviasi yang tidak dapat dijelaskan sepenuhnya oleh model linear.

REGRESI DENGAN MODEL PANGKAT SEDERHANA

Regresi dengan Model Pangkat Sederhana (Power Law Regression) adalah teknik pemodelan statistik yang digunakan untuk menggambarkan hubungan antara dua variabel yang mengikuti pola fungsi pangkat. Ini berbeda dari regresi linear dalam bentuk matematis yang digunakan untuk menggambarkan hubungan antara variabel-variabel tersebut. Dalam model pangkat sederhana, hubungan antara variabel independen X dan variabel dependen Y dinyatakan dalam bentuk:

$$Y = aX^b$$

Di mana:

Y adalah variabel dependen.

X adalah variabel independen.

a adalah koefisien skala.

b adalah eksponen yang menunjukkan tingkat hubungan non-linear antara X dan Y .

Implementasi Kode

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error
from scipy.optimize import curve_fit

url = 'C:/Users/gary/python/Numerik-
Regresi/student_performance.csv'
data = pd.read_csv(url)
print(data.head())

X = data['Hours Studied'].values.reshape(-1, 1)
y = data['Performance Index'].values
```

```

def power_law(X, a, b):
    return a * np.power(X, b)

popt, pcov = curve_fit(power_law, X.flatten(), y)

y_pred_power = power_law(X.flatten(), *popt)

rms_error_power = np.sqrt(mean_squared_error(y, y_pred_power))

plt.scatter(X, y, color='blue', label='Data Asli')
plt.plot(X, y_pred_power, color='red', label='Regresi Pangkat
Sederhana')
plt.xlabel('Hours Studied')
plt.ylabel('Performance Index')
plt.title('Regresi Pangkat Sederhana')
plt.legend()
plt.show()

print('RMS Error untuk Model Pangkat Sederhana:',
rms_error_power)

```

Ringkasan

Kode di atas melakukan analisis regresi pangkat sederhana pada data yang terdiri dari dua variabel: 'Hours Studied' (jam belajar) dan 'Performance Index' (nilai ujian). Tujuan utamanya adalah memodelkan hubungan antara jam belajar dan nilai ujian menggunakan regresi pangkat sederhana dan mengevaluasi kesalahan model menggunakan Root Mean Squared Error (RMSE).

Konsep

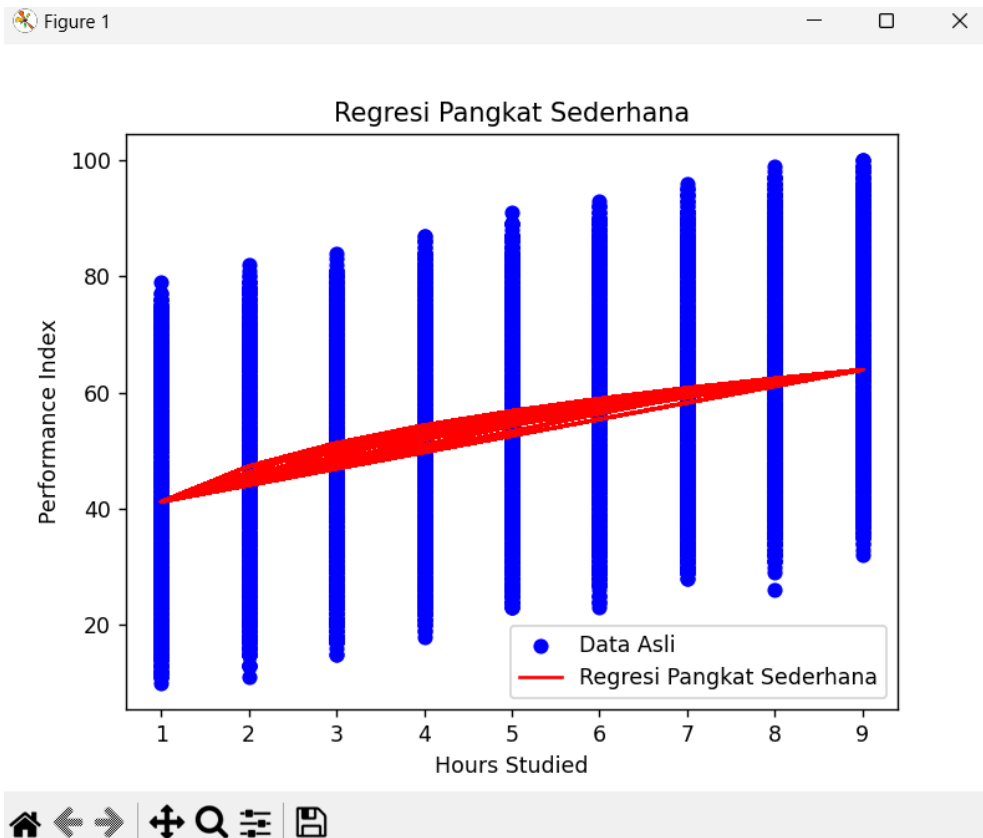
1. Regresi Pangkat: Teknik statistik yang memodelkan hubungan antara variabel dependen (y) dan variabel independen (X) dalam bentuk fungsi pangkat. Model yang digunakan adalah $y = a \cdot X^b$, di mana a dan b adalah parameter model.

2. Curve Fitting: Menggunakan metode `curve_fit` dari SciPy untuk menyesuaikan model pangkat dengan data yang diamati.
3. Root Mean Squared Error (RMSE): Ukuran kesalahan antara nilai yang diprediksi oleh model dan nilai sebenarnya. RMSE memberikan gambaran tentang seberapa baik model memprediksi data yang sebenarnya.
4. Visualisasi: Menggunakan scatter plot untuk menampilkan data asli dan garis regresi pangkat untuk menunjukkan hasil model pangkat.

Hasil Pengujian

```
C:\Users\gary\python>C:/Users/gary/AppData/Local/Programs/Python/Python311/
python.exe c:/Users/gary/python/Numerik-Regresi/PangkatSederhana.py
Hours Studied ... Performance Index
0          7 ...          91.0
1          4 ...          65.0
2          8 ...          45.0
3          5 ...          36.0
4          7 ...          66.0

[5 rows x 6 columns]
RMS Error untuk Model Pangkat Sederhana: 17.886378846062684
```



Analisis Hasil

1. Hubungan Non-Linear: Scatter plot dan garis regresi pangkat menunjukkan bahwa ada hubungan non-linear antara 'Hours Studied' dan 'Performance Index'. Hal ini berarti hubungan antara kedua variabel tidak linier, dan model pangkat sederhana lebih sesuai dalam menangkap pola data.
2. Kesalahan Model: RMSE memberikan gambaran tentang seberapa baik model melakukan prediksi. Nilai RMSE yang lebih rendah dibandingkan dengan model linear menunjukkan bahwa model pangkat mungkin lebih akurat dalam memprediksi nilai kinerja berdasarkan jam belajar.
3. Kekuatan Model: Garis regresi yang digambarkan pada plot menunjukkan bahwa model pangkat cukup baik dalam menangkap tren umum dalam data, dengan lebih sedikit deviasi atau outlier dibandingkan model linear.

Berdasarkan analisis dari kedua model di atas, model regresi pangkat sederhana lebih baik dalam menangkap hubungan antara Hours Studied dan Performance Index, memberikan prediksi yang lebih akurat dan representatif daripada model regresi linear.